# Ontologie, sciences cognitives et identité personnelle :

Mikaël Mugneret

▶ **To cite this version:**

Mikaël Mugneret. Ontologie, sciences cognitives et identité personnelle :. Philosophie. Université Nancy 2, 2006. Français. NNT : 2006NAN21012 . tel-01751193

**HAL Id: tel-01751193**
**https://hal.univ-lorraine.fr/tel-01751193**

Submitted on 29 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Université Nancy 2

LPHS - Archives Henri Poincaré, CNRS UMR 7117

Ecole doctorale Langues, Temps, Sociétés

Maison des Sciences de l'Homme Lorraine

# Ontologie, sciences cognitives et identité personnelle

Mikaël MUGNERET

Thèse présentée pour l'obtention du
Doctorat de Philosophie de l'Université Nancy 2

Sous la direction de Monsieur le Professeur Roger POUIVET
(Archives Poincaré, CNRS, Université Nancy 2)

Membres du jury :

M. Gerhard HEINZMANN
Professeur, Archives Poincaré CNRS, Université Nancy 2

Mme Anne GIERSCH-LUN
Chargée de Recherche, INSERM, Université Louis Pasteur, Strasbourg 1

M. Pierre LIVET
Professeur, CEPERC CNRS, Université de Provence, Aix-Marseille 1

M. Jérôme DOKIC
Directeur d'Etudes, École des Hautes Études en Sciences Sociales, Paris

M. Roger POUIVET
Professeur, Archives Poincaré CNRS, Université Nancy 2

Année universitaire 2006-2007

# Remerciements :

Je tiens tout d'abord à remercier le Professeur Roger Pouivet d'avoir accepté de diriger ce travail jusqu'à son aboutissement. Sa disponibilité, son soutien, ses encouragements et ses conseils, tout au long de ces quatre années de D.E.A puis de doctorat, ont été une aide très précieuse à l'élaboration de cette thèse.

Je voudrais également remercier Manuel Rebuschi. Les enseignements que j'ai reçus de lui lors de ma licence et maîtrise de sciences cognitives ont suscité mon intérêt pour la philosophie de l'esprit. C'est également à lui que je dois d'avoir rencontré le Professeur Roger Pouivet.

Merci à toute l'équipe des « Archives Poincaré », et à son directeur, le Professeur Gerhard Heinzmann, pour son accueil et sa sympathie.

Merci à Anne Giersch-Lun, Jérôme Dokic et au Professeur Pierre Livet, pour avoir accepté de faire partie de mon jury de thèse, et d'avoir fait le déplacement jusqu'ici.

Merci à mes ami(e)s, pour leur appui, leur intérêt et leurs encouragements.

A mes parents, Jeanine et Christian Mugneret, ainsi qu'à mon parrain, Jean Zehr, un grand merci d'être là, de m'avoir motivé à faire des études supérieures, et de m'avoir donné les moyens matériels, intellectuels et affectifs de les réaliser. Un grand merci aussi pour votre relecture, et pour l'intérêt manifesté à cette thèse.

A ma future belle-famille, Margaret et Virginie Sprunck, un grand merci aussi, pour votre soutien, votre relecture de ce travail, et pour l'intérêt porté à cette thèse.

Enfin, je voudrais remercier ma fiancée, Catherine Sprunck, à laquelle je dédie cette thèse, pour son amour, sa patience, son soutien, ses relectures et ses encouragements. Merci à toi, pour l'intérêt que tu as porté à cette thèse. Merci aussi, pour les discussions enrichissantes que nous avons, et qui en ont influencées l'écriture et le contenu, dans une mesure que je ne saurais évaluer.

“My life is an endless succession of people saying goodbye
My life is an endless succession of people saying goodbye
And what’s left for me?
What’s left for me?

At one time the future it stretched out before me
But now it stretches behind me.
And all of the best things in life are behind glass
Money, jewellery and flesh
And what’s left for me?
What’s left for me?”

(Morrissey, “My Life is a Series of People Saying Goodbye”)

# Table des matières :

**PARTIE 1 : INTRODUCTION 11**

**I. INTRODUCTION 15**
**II. METHODOLOGIE 25**
A. METAPHYSIQUE ET ONTOLOGIE, SCIENCES ET EPISTEMOLOGIE. 25
B. LA METHODOLOGIE DES EXPERIENCES DE PENSEE 36
1. De la nature des expériences de pensée 37
2. Fonctionnement et épistémologie des expériences de pensée 38
3. Types d'expériences de pensée 41
**III. ELEMENTS DE GEOGRAPHIE CONCEPTUELLE 44**
A. ESPACE DU PROBLEME 44
1. Identité personnelle versus identité individuelle 44
2. Trois axes au problème 46
3. Pourquoi est-ce bien moi qui suis moi ? 48
4. Problèmes de l'unité expérientielle et consciente 50
5. Fondements de l'identité numérique : (1) identité qualitative, ressemblance, connexité ; (2) continuité (qualitative et spatiale) et (3) identité numérique des constituants. 61
B. ESPACE DES SOLUTIONS 63
1. Réductionnisme vs Antiréductionnisme 65
2. Biologisme vs Psychologisme 68
3. Internalisme vs externalisme 73

**PARTIE 2 : DISCUSSION ET CRITIQUE 75**

**I. INTRODUCTION 79**
**II. LES INTUITIONS ANTIREDUCTIONNISTES SUR L'IDENTITE PERSONNELLE 93**
A. INTUITION N°1 : LE FAIT MANQUANT. 98
B. INTUITION N°2 : LE « MOI » FIXE ET PERMANENT. 100
C. INTUITION N° 3 : LE « MOI » INTERNE. 101
D. INTUITION N°4 : J'AURAIS PU ETRE N'IMPORTE QUI D'AUTRE. 102
E. INTUITION N°5 : LES DESIGNATEURS RIGIDES. 115
F. INTUITION N°6 : LES APPROCHES EN 1ERE ET 3EME PERSONNE SONT IRREDUCTIBLES. CHAQUE APPROCHE EN 1ERE PERSONNE EST IRREDUCTIBLE A UNE AUTRE APPROCHE EN 1ERE PERSONNE. 118
G. INTUITION N°7 : ETRE UNE PERSONNE, C'EST ETRE UNE PERSONNE. 124
H. INTUITION N° 8 : L'INDEXICAL ESSENTIEL. 130
I. INTUITION N°9 : L'IRREDUCTIBLE LIBRE-ARBITRE 132
J. INTUITION N°10 : LA DESINCARNATION PROGRESSIVE 153
K. INTUITION N°11 : LES ZOMBIES 154
L. INTUITION N°12 : LE NEANT SUBJECTIF INIMAGINABLE 155
M. CONCLUSION 156
**III. DES CRITERES APORETIQUES 165**
A. CRITERE PSYCHOLOGIQUE : 165
1. Circularité : 165
2. Non-transitivité 167
3. Duplication : 170
B. CRITERE SOMATIQUE 200

1. Le critère somatique est-il intuitif ? 205
2. L'amnésie antérograde 207
3. Critère somatique et présent spécieux 213
4. La meilleure image de l'âme humaine. 224
**IV. DE L'INTERNALISME A L'EXTERNALISME 228**
A. INTRODUCTION 228
B. PERSPECTIVE EPISTEMOLOGIQUE 239
1. L'argument du langage privé et ses variantes 239
2. La connaissance par contact direct : réalité et limites 247
3. Réhabilitation d'une forme de phénoménalisme 273
C. PERSPECTIVE ONTOLOGIQUE 282
1. L'esprit étendu 290
2. Psychologie in vitro 293
3. De l'illusion d'habiter son corps 299
4. Cerveau planétaire et sociologie neuronale 306
5. Conclusion 308
D. APPLICATION DE L'EXTERNALISME AU PROBLEME DE L'UNITE DE LA CONSCIENCE 313
1. Les patients split brain 315
2. Les jumeaux « siamois » 341

**PARTIE 3 : IDENTITE PERSONNELLE ET EXPERIENCES DE PENSEE 345**

**I. INTRODUCTION 349**
**II. DU CONCEVABLE AU POSSIBLE 361**
A. DIFFERENTES MANIERES DE CONCEVOIR LA CONCEVABILITE 362
B. FOSSES ENTRE LA CONCEVABILITE ET LA POSSIBILITE : 366
C. DE LA CONCEVABILITE A LA POSSIBILITE : L'ARGUMENT « PRAGMATIQUE » DE CHALMERS 367
D. DE LA CONCEVABILITE A LA POSSIBILITE : UN ARGUMENT RATIONALISTE 372
**III. EXPERIENCES DE PENSEE ET IDENTITE PERSONNELLE 378**
A. LE PHILOSOPHE ET SON CANTILEVER… 381
1. L'argument de Gendler 382
2. Le soi et le futur 385
3. La morale des histoires 388
4. Commentaires 392
B. DES DESACCORDS SUR LA METHODE DE L'ACCORD 410
1. L'argument de la fission selon Parfit : 412
2. Pourquoi l'argument de la fission paraît si convaincant ? 415
3. Objections à Gendler 420
C. BILAN SUR GENDLER 432
**IV. METAPHYSIQUE DU « MOI » : UTILITE ET FONDEMENT. 435**
A. A QUOI ÇA SERT ? 435
B. Y A-T-IL UN « MOI » REEL, AVEC DES CONDITIONS D'IDENTITE A TRAVERS LE TEMPS ? 442
1. Réalisme ou anti-réalisme ? 442
2. Le réalisme du « moi » 447
C. QUEL EST DONC CE « MOI » DONT ON RECHERCHE LES CONDITIONS D'IDENTITE A TRAVERS LE TEMPS ? 451

**PARTIE 4 : CONCLUSION 465**

**ANNEXE : IDENTITE PERSONNELLE ET SURVIE 487**

**I. INTRODUCTION 491**
**II. CRITIQUE DE LA NOTION DE NEANT 496**
A. L'APPROCHE POSITIVE DU NEANT (LE NEANT COMME ETAT DURANT LEQUEL NOUS N'EXPERIMENTONS RIEN) ET L'ARGUMENT LOGIQUE : UN NEANT POSITIF EST LOGIQUEMENT CONTRADICTOIRE. 498
B. L'APPROCHE NEGATIVE DU NEANT (LE NEANT N'EST PAS UN ETAT) ET L'ARGUMENT SEMANTIQUE/NON-COGNITIVISTE : LE CONCEPT D'ANNIHILATION EST UN CONCEPT SEMANTIQUEMENT VIDE. 502
C. CONCLUSION 504
**III. ARGUMENTS POUR L'ETERNITE 509**
A. IL ETAIT A PRIORI TRES IMPROBABLE QUE J'EXISTE MAINTENANT 509
B. CE QUI NE DURE PAS NE PEUT PAS ETRE DEFINITIF 513
C. LE TEMPS NE S'ECOULE PAS REELLEMENT 515
D. UN ARGUMENT GENERAL EN FAVEUR DE L'ETERNITE DU SUJET. 519
**IV. QUELLE IDENTITE PERSONNELLE POUR QUELLE SURVIE ? 522**
A. PREMIER MODELE : THEORIE PHENOMENALISTE DE L'IDENTITE PERSONNELLE. 523
B. DEUXIEME MODELE : THEORIE REALISTE ONTOLOGIQUE DE L'IDENTITE PERSONNELLE. 531
C. TROISIEME MODELE : THEORIE PHYSICALISTE INTEGRATIVE DE L'IDENTITE PERSONNELLE. 534
D. SYNTHESE. 542

**BIBLIOGRAPHIE ET INDEX 547**

# Partie 1 : Introduction

« Est-ce que, par hasard, on m'aurait changée au cours de la nuit ? Réfléchissons : *étais*-je identique à moi-même lorsque je me suis levée ce matin ? Je crois bien me rappeler m'être sentie un peu différente de l'Alice d'hier. Mais, si je ne suis pas la même, il faut se demander alors *qui* je peux bien être ? Ah, c'est *là* le grand problème ! »

(Lewis Carroll, *Les Aventures d'Alice au Pays des Merveilles*)

# I. Introduction

« Tu vois, là, dans le landau, **c'est moi quand j'avais** 5 mois. »

« Et là, c'est mon frère jumeau. On est **tellement semblables** que tout le monde nous confond »

« Ah, quant à celui-là, c'était toujours le premier de la classe ! Comme **j'**aurais bien voulu **être lui** ! »

Des discours de ce genre, la plupart d'entre nous en avons déjà entendu souvent, à l'occasion, par exemple, de la consultation d'un vieil album photo. Il se dégage de ces expressions un ensemble de présupposés sur l'identité personnelle. Des présupposés que nous qualifierons d'antiréductionnistes, et que nous pouvons résumer en quelques points :

1. **La singularité :** Je suis une entité singulière, unique. En revanche, je ne suis pas les autres, je suis un être singulier, séparé, indépendant ontologiquement des autres. A la limite, il pourrait être concevable que je sois seul à exister et que tous les autres soient illusion. Mon frère jumeau, bien que me ressemblant énormément, n'est pas moi. A la limite, il pourrait être concevable que je sois seul à exister et que tous les autres ne soient que des « zombis » ou de pures illusions, produits de mon esprit. Tandis que moi, je ne peux qu'exister, car pendant que je douterais que j'existe, je ne pourrais pas douter que je doute, et donc que je pense, car le doute est une forme de pensée. Et du fait même que je

pense, il s'ensuit que j'existe car pour penser, il faut déjà exister. Je ne peux être absolument sûr que de ma propre existence.

2. **La persistance :** Je suis une entité persistante. Je suis le même « Je » depuis ma naissance jusqu'à ma mort, en dépit des changements qui peuvent m'affecter. Je suis l'enfant que j'étais. Je suis le vieillard que je serai. Tout au long de mon existence, je peux énormément changer. Il semble même que je peux davantage ressembler à mon frère jumeau que je ne me ressemble à moi-même deux années de suite. Il semble qu'il n'y a pas de limite aux changements que je puis subir : je peux perdre ma mémoire, changer radicalement de personnalité, passer d'un bébé joufflu à un vieillard sénile, voire le stock de molécules qui me compose être entièrement renouvelé, sans que quelqu'un meurt, disparaisse, et que quelqu'un d'autre naisse, apparaisse à la place. Le seul lien entre ces états successifs semble être une certaine continuité : continuité corporelle, continuité dans l'espace, continuité psychologique, continuité mnésique. Mais ce lien paraît trop faible face à la force de ce qui unit tous les instants de ma vie. Il est nécessaire que je sois cet enfant, représenté sur cette photo. Mais était-il nécessaire que je devienne qui je suis aujourd'hui ? N'aurait-il pas pu m'arriver autre chose ? N'aurais-je pas pu faire d'autres choix ? Et quelle force doit avoir cette soi-disant continuité pour fonder ma pérennité ? Aurait-il suffit d'un tout petit défaut dans cette dernière pour que je n'existe pas aujourd'hui ?

3. **L'indivisibilité :** Je suis indivisible, en vertu du fait que je suis unique. Si je pouvais être divisé en deux moitiés de moi-même séparées, elles deviendraient, l'une pour l'autre, des étrangères. Chacune pourrait douter de la réelle existence

de l'autre et pourrait se croire seule au monde. Je ne peux pas expérimenter simultanément deux contenus de conscience dissociés tout en expérimentant cette dissociation.

4. **L'ineffabilité :** J'aurais pu être quelqu'un d'autre de complètement différent. J'aurais pu être Benoît XVI par exemple. Inversement, quelqu'un d'autre que moi aurait pu exister, qui aurait été en tout point indiscernable de moi, mais qui n'aurait pas été moi. A ce sujet, il existe une plaisanterie philosophique fameuse et déconcertante : « Saviez-vous que Shakespeare n'a jamais existé et que toutes ses œuvres ont en fait été écrites par quelqu'un d'autre portant le même nom ? ». Je ne suis donc pas constitué de matière ou de traits psychologiques.

5. **La liberté :** Je suis doué de libre-arbitre, et donc de responsabilité morale. Je ne suis pas entièrement déterminé par les gènes et l'environnement. Je suis en mesure de m'autodéterminer moi-même et de faire de bons ou de mauvais choix. La culpabilité a un sens. Le mérite également.

6. **La conscience phénoménale :** J'expérimente certaines choses en moi-même que je suis le seul à expérimenter : une douleur au pied, la faim, un mal de dent, etc. Bien sûr, d'autres personnes peuvent expérimenter la faim, par exemple. Mais elles ne peuvent pas expérimenter *ma* faim, de même que je ne peux pas expérimenter *leur* faim. Elles ne peuvent pas avoir faim à ma place et je ne peux pas avoir faim à leur place.

7. **L'immortalité, voire l'éternité :** Je ne peux pas m'imaginer comme non existant. Si j'essaye, je me rends compte qu'il en reste toujours quelque chose : quelqu'un essayant de s'imaginer non existant ! Le plus parfait état d'isolation sensoriel et de retenue de la pensée n'équivaut pas à un pur néant, bien au contraire. J'en conclus donc à mon immortalité, de par l'impossibilité de m'imaginer, en première personne, annihilé. J'en conclus aussi à mon immatérialité, la matière de mon corps étant évidemment vouée à la corruption.

De manière un peu plus littéraire, nous ne résistons pas à l'envie de vous faire partager ce texte de Douglas Harding, qui exprime, avec une touchante naïveté, un certain retour aux sources de ces intuitions antiréductionnistes, sous leur forme la plus pure, débarrassée des rationalisations de la science et de la philosophie :

> « Le plus beau jour de ma vie – ma nouvelle naissance en quelque sorte – fut le jour où je découvris que je n'avais pas de tête. Ceci n'est pas un jeu de mots, une boutade pour susciter l'intérêt coûte que coûte Je l'entends tout à fait sérieusement : je n'ai pas de tête. Je découvris instantanément que ce rien où aurait dû se trouver une tête, n'était pas une vacuité ordinaire, un simple néant. Au contraire, ce vide était très habité. C'était un vide énorme, rempli à profusion, un vide qui faisait place à tout – au gazon, aux arbres, aux lointaines collines ombragées et, bien au-delà d'elles, aux cimes enneigées semblables à une rangée de nuages anguleux parcourant le bleu du ciel. J'avais perdu une tête et gagné un monde.
>
> Tout cela me coupait littéralement le souffle. Il me semblait d'ailleurs que j'avais cessé de respirer, absorbé par Ce-qui-m'était-donné : ce paysage superbe, intensément rayonnant dans la clarté de l'air, solitaire sans soutien, mystérieusement suspendu dans le vide, et (en cela résidait le vrai miracle, la merveille et le ravissement) totalement exempt de

> « moi », indépendant de tout observateur. Sa présence totale était mon absence totale de corps et d'esprit »[1]

Une enquête empirique et rationnelle ne peut pas justifier ou servir d'assise à nos intuitions antiréductionnistes sur l'identité personnelle. Nous le montrerons dans ce travail. Pourtant, une fois mis à plat et décortiqués les faux raisonnements qu'on invoque à l'appui de celles-ci, on reste sur un sentiment d'insatisfaction, comme si on était passé à côté de quelque chose d'essentiel, comme si, par mégarde, on avait jeté le bébé avec l'eau du bain, et pourtant, on a beau cherché, on ne retrouve pas le bébé ! Les intuitions antiréductionnistes sur l'identité personnelle semblent très prégnantes, indéracinables, même quand on en a compris – ou croit en avoir compris – les mécanismes psychologiques.

Ce qui est proprement étonnant, c'est que ce sentiment persiste, quand bien même on serait intimement convaincu du caractère fallacieux des arguments généralement avancés à l'appui d'une conception antiréductionniste de l'identité personnelle et du bien-fondé des arguments généralement avancés à l'appui d'une conception réductionniste de l'identité personnelle.

On est un peu dans la situation bien connue, suivante : Lorsqu'un arbre tombe, nous pouvons constater, lorsque nous sommes à proximité, qu'il fait du bruit en tombant. Mais fait-il aussi du bruit en tombant lorsque personne n'est présent pour l'écouter ? On est tenté de dire oui, et pourtant, que serait un bruit qui ne serait entendu par personne ?

---

[1] HARDING, Douglas. « Vivre sans Tête ». *In* : DENNETT, Daniel C., et HOFSTADTER, Douglas. *Vues de l'Esprit*, p. 32.

Malgré cette explication, on aura du mal à admettre que les arbres ne font pas du bruit en tombant, lorsqu'il n'y a personne pour entendre ce bruit.

L'exemple de l'arbre est peut-être plus proche de nos questionnements sur l'identité personnelle qu'on pourrait le penser.

En effet, il s'inscrit dans un débat qui oppose les tenants d'une option réaliste, en métaphysique, des tenants d'une option antiréaliste.

Pour les réalistes, il existe un monde indépendant de nos facultés de le connaître. Pour les antiréalistes, un tel monde n'existe pas ou du moins on ne peut affirmer légitimement son existence. Les réalistes s'appuient souvent sur une inférence à la meilleure explication : l'existence des entités physiques inobservables serait la meilleure explication à l'efficacité de nos théories physiques.

Les antiréalistes s'appuient sur le rasoir d'Occam, qui recommande de ne pas multiplier les entités sans nécessité. Seul ce qui peut être directement observé peut légitimement être dit exister.

Une analyse strictement objective du monde, ne faisant appel qu'à l'observable (voire, pour certains extrémistes, à l'observé actuellement) et à la logique, aboutira normalement à une position antiréaliste. C'est ainsi que Ernst Mach, ce fameux philosophe précurseur du non moins fameux Cercle de Vienne, nia l'existence des atomes, car ces derniers ne pouvaient être traduits en énoncés protocolaires, c'est-à-dire en énoncés d'observation directe. Contre cette conception, Wittgenstein argumentera que la croyance

précède le doute, qu'il existe certaines propositions de base que nous sommes obligés d'admettre, même si nous ne pouvons les justifier. Que la justification, de toute manière, doit nécessairement s'arrêter à un moment[2].

L'hypothèse de l'âme, comme soubassement inobservable de l'identité personnelle, peut-elle être maintenue à titre d'intuition métaphysique commode et consensuelle ; au même titre que les forces physiques, les atomes, les quarks, les nucléons, les électrons, les photons, les neutrinos, etc., dont l'existence n'est généralement pas considérée comme problématique par la plupart des scientifiques non formés aux subtilités de la philosophie antiréaliste ? Nous verrons que rien ne nous en empêche. Par effet de symétrie, si l'on doit être réaliste au sujet des entités physiques, alors nous sommes fondés à l'être pour d'autres choses, dont l'identité personnelle.

Toutefois, un travail philosophique digne de ce nom ne saurait en rester là. Et quand bien même nous serions justifiés à croire qu'une approche réductionniste de l'identité personnelle ne fonde pas, ontologiquement, cette dernière, on peut néanmoins maintenir qu'elle l'explique, sur un plan épistémologique. Ainsi, pour prendre une analogie avec la religion : on peut ne pas croire en Dieu et accepter l'idée d'une évolution néodarwinienne des êtres vivants ; on peut croire en Dieu[3] et rejeter toute idée d'évolution néodarwinienne des êtres vivants, Dieu étant le créateur direct et efficient de toute forme de vie ; mais on peut aussi croire en Dieu et penser qu'il a créé les êtres vivants, par le biais de l'évolution néodarwinienne des êtres vivants… Dans les deux premiers cas, le conflit vient de ce que l'on tient à placer les deux types d'explications de l'origine des êtres vivants sur le même

---

2 WITTGENSTEIN, Ludwig. *De la Certitude*.

3 Ou croire en un Dieu qui n'a pas créé les êtres vivants.

plan (ontologique et/ou épistémologique). Mais la troisième option est valide également, et se comprend dans la mesure où elle prend soin de distinguer ce qui relève d'une explication scientifique de l'origine des êtres vivants, de ce qui relève d'une explication métaphysique, ontologique ou théologique de cette origine. Les deux explications ne se contredisent pas car elles ne sont pas du même type. L'explication scientifique nous donne le processus effectif (cause efficiente ou motrice) par lequel les êtres vivants sont apparus, tandis que l'explication métaphysico-ontologico-théologique inscrit ce processus dans une dimension téléologique (cause finale). Dit autrement, la science élucide les « comment », décrit logiquement les chaînes de causes et d'effets objectivement observables ; tandis que la métaphysique, l'ontologie et la théologie proposent des « pourquoi », donnent des raisons, confèrent un sens, à ce qui est observable, qui toutefois dépasse ce dont on est en mesure de pouvoir observer.

De même, une explication antiréductionniste de l'identité personnelle n'est pas nécessairement incompatible avec une explication réductionniste. Il suffit que ces deux explications ne soient pas au même niveau. Une explication réductionniste de l'identité personnelle nous donnerait le processus effectif par lequel se constitue notre sentiment de nous-mêmes et des autres, proposerait des critères utilisables en pratique pour identifier les personnes, tandis qu'une explication antiréductionniste inscrirait tout cela dans la perspective plus large d'une appréhension subjective de notre identité. Le challenge sera de faire en sorte de justement réussir à concilier ces deux approches, de les rendre au plus compatibles l'une avec l'autre.

Autour de la notion d'identité personnelle gravite toute une série de problématiques connexes et enchevêtrées au nombre desquelles on peut citer : la question de l'identité

personnelle à travers le temps, l'espace, les mondes possibles, la distinction entre individu et personne, et donc entre identité individuelle et identité personnelle *stricto sensu*, le problème de l'unité synchronique et diachronique de la conscience et de l'expérience, la distinction entre divers types d'identité (numérique [absolue ou relative], sortale, qualitative), entre connexité (*connectedness*) et continuité (*continuity*), etc. Après avoir passé en revue l'espace du problème lié à l'identité personnelle, nous nous concentrerons plus particulièrement sur la question de l'identité personnelle à travers le temps.

Nous opposerons tout d'abord les perspectives radicales incarnées par l'antiréductionnisme et le réductionnisme, notamment éliminativiste. Nos intuitions concernant l'identité personnelle résistent-elles à l'examen ? Peuvent-elles espérer être étayées par une argumentation rationnelle ?

Puis, dans une perspective réductionniste non-éliminativiste, nous examinerons les critères classiques rencontrés dans la littérature : critères psychologiques et somatiques seront tout d'abord exposés, puis critiqués. Tous ces critères sont scientifiquement acceptables voire plausibles, mais ne sont-ils pas contre-intuitifs dans certains cas ? La recherche de bons critères passera donc surtout par la mise en balance de nos intuitions. Cela fait toute l'originalité de ce genre d'investigation, essentiellement conceptuel dans sa forme, qui renoue sur le fond avec l'ontologie (au sens d'examen de nos présomptions sur le réel), et du mode d'argumentation, qui s'appuie le plus souvent sur des expériences de pensée visant à remettre en question nos intuitions ontologiques sur l'identité personnelle. Pour être plausible, cette réflexion ontologique, pour conceptuelle qu'elle soit, ne doit pas nier les faits. Elle doit, au contraire, s'efforcer de rendre compte au mieux des résultats les plus aboutis obtenus en sciences cognitives notamment. Nous verrons si, malgré cette

double exigence contradictoire – respecter au mieux *les faits* ainsi que *nos intuitions ontologiques* – une sauvegarde du concept d'identité personnelle en philosophie de l'esprit est possible.

Ensuite, quelques aspects, en philosophie de l'esprit, de la controverse entre internalisme et externalisme, seront évoqués. Nous en explorerons quelques applications possibles par rapport à la question de l'identité personnelle, ainsi que de l'unité de la conscience et de l'expérience.

Enfin, nous approfondirons la méthodologie des expériences de pensée, souvent employée dans le contexte des réflexions sur l'identité personnelle. Cette méthode est-elle vraiment fiable ? Tout d'abord, la question se posera de l'utilité voire de la légitimité philosophique à rechercher des critères de l'identité personnelle fonctionnant dans tous les cas *possibles* (scénarios type « science-fiction ») voire même simplement dans les cas réels *marginaux* (patients split-brain, amnésies sévères, etc.). Nous montrerons que la réponse à cette question dépend du type d'enquête que l'on souhaite mener (ontologique ou seulement conceptuelle) et/ou de l'ontologie à laquelle on adhère (réalisme ou instrumentalisme).

Dans les annexes, les personnes intéressées par les controverses liées à l'après-vie pourront lire quelques réflexions à ce sujet. La question de la survie et celle de l'identité personnelle sont historiquement et logiquement liées. Elles peuvent s'apporter un éclairage réciproque, comme nous espérons réussir à le montrer.

# II. Méthodologie

## *A. Métaphysique et ontologie, sciences et épistémologie.*

Croyez-vous que Dieu existe ?

A cette question, certains répondent parfois : « Dites-moi ce que vous entendez par Dieu, et je vous dirai si j'y crois ».

Mais on pourrait aussi répondre ceci :

« Dites-moi ce que vous entendez par *croire*, et je vous dirai si j'y crois ». Croire est-ce seulement répondre « je le crois » à la question « est-ce que *p* est vrai ? » ou à la question « est-ce que *A* existe ? ». N'est-ce pas aussi et même davantage un certain comportement, un certain nombre d'images mentales, d'émotions ? Comme le souligne Roger Pouivet[4] :

> « [U]ne croyance est essentiellement dispositionnelle. Croire en Dieu, c'est, par exemple, avoir tendance à prier en certaines circonstances, au moins dans les moments difficiles ou dans les moments particulièrement heureux de l'existence, se sentir alors aidé, soutenu, avoir le désir de remercier, avoir mauvaise conscience quand on agit d'une certaine façon, etc. »

Un tel comportement pourrait très bien, jusqu'à un certain point, être compatible avec une affirmation explicite d'athéisme – au moins dans sa version étymologiquement correcte d'*absence de croyance* en Dieu et non pas de *croyance en son absence* – de la même

---

[4] POUIVET, Roger. « Epistémologie de la croyance religieuse ». *In :* BOURGEOIS-GIRONDE, Sacha, GNASSOUNOU, Bruno, et POUIVET, Roger, *Analyse et théologie : Croyances religieuses et rationalité*, p. 23.

manière où, par exemple, certaines personnes ont une peur viscérale de prendre l'avion, tout en reconnaissant aisément que d'un point de vue rationnel, il est peu probable qu'un avion s'écrase. Mais sait-on jamais ?...

La croyance en Dieu, dans certains cas au moins, peut-elle avoir un statut de connaissance ? On remarque que certaines personnes, notamment celles ayant vécu des expériences dites mystiques, mais aussi, dans une moindre mesure, celles prétendant fonder leur croyance en Dieu sur des bases rationnelles et empiriques, disent parfois des choses du genre : « Dieu, pour moi, ce n'est pas une question de croyance, c'est une certitude, un savoir, une connaissance, un fait indubitable. ». Or, une connaissance est souvent considérée comme une croyance *vraie* et *justifiée*. Répondre à la question implique donc de se pencher sur les concepts de *vérité* et de *justification*.

A la question posée en en-tête, on pourrait également répondre : « Dites-moi ce que vous entendez par *exister*, et je vous dirai si j'y crois ». En effet, Dieu, s'il existe, n'est certainement pas censé exister de la même manière dont existent les êtres humains[5], les animaux, les plantes, les roches, les artefacts, etc. Nul théologien n'a jamais prétendu avoir aperçu Dieu quelque part au coin d'une rue ou au détour d'un chemin. Aucun n'a jamais eu

[5] Le cas du mormonisme (Eglise de Jésus-Christ des Saints des Derniers Jours) constitue une remarquable exception, à cet égard.

l'audace de l'établir suite à une expérience de laboratoire[6]. Mais si Dieu n'existe pas de cette manière là, dont existent les êtres usuels, de quelle manière peut-il être dit exister ?

Répondre correctement à la question : « Croyez-vous que Dieu existe ? », nécessite donc une approche croisée entre diverses approches : théologie, ontologie et épistémologie.

L'épistémologie s'intéresse à définir normativement ce qu'est une connaissance, c'est-à-dire à cerner quelles sont les conditions nécessaires et suffisantes pour qu'une proposition donnée soit considérée comme une connaissance, par opposition, par exemple, à une simple croyance. Il s'agira donc d'élucider le concept de connaissance, de dégager les principes qui sous-tendent toute connaissance possible. Par extension, l'épistémologie s'intéresse aussi aux croyances, selon une modalité semblable. Il existe ainsi une épistémologie de la croyance religieuse. Sans doute conviendrait-il de distinguer entre une épistémologie véritable (*épistémè* signifie connaissance) étude philosophique de la connaissance, et ce que l'on pourrait appeler une doxologie (*doxa* étant l'opinion). L'épistémologie est très liée aux sciences, qui en constituent l'application concrète visant à acquérir des connaissances positives, et surtout aux sciences cognitives, qui se penchent sur les processus effectifs par lesquels les êtres connaissants acquièrent, traitent, emmagasinent, utilisent, et justifient leurs croyances et connaissances.

Alors que l'épistémologie s'intéresse à la connaissance, l'ontologie s'attache à décrire comment doit être le monde pour qu'il soit connaissable. Précisons tout de suite que

---

[6] A la regrettable exception des partisans de la *Creation Science* ou de l'*Intelligent Design* aux USA, et des membres de l'Université Interdisciplinaire de Paris (UIP) en France. Bien qu'elles se revendiquent scientifiques, ces approches violent cependant les principes et méthodes de la science. Elles constituent donc des pseudosciences.

l'ontologie n'a pas nécessairement vocation à dire comment le monde est réellement, mais uniquement à dire comment il devrait être pour être intelligible et objet de connaissance. Il se peut que le monde soit autrement que ce que nous dit l'ontologie, mais alors il est au moins en partie inconnaissable voire inconcevable. L'ontologie s'interroge donc sur le concept d'existence, soit de manière générale : « Qu'est-ce que le réel ? Qu'est-ce que l'être ? »[7], ou encore : « Qu'est-ce qui est réel ? Qu'est-ce qui est ? »[8], soit de manière plus spécifique : « De quelle manière existe tel ou tel type d'être ? », ou encore : « De quels sous-types est constitué tel ou tel type d'être ? » (en termes wittgensteiniens : « Quelle est la grammaire de ce concept ? »).

Enfin, l'on peut s'interroger ontologiquement sur des êtres individuels et non pas sur des types d'êtres : « Quelle est l'essence de cet être là ? » ou encore : « De quels êtres individuels est constitué le monde ? » (Par exemple, existe-t-il un Dieu ou une Déesse, des anges, des démons ? L'être humain a-t-il une âme immortelle ?). Ces derniers types d'interrogation sont cependant plutôt rangés dans le domaine de la métaphysique. Ces êtres individuels ne sont pas nécessairement des particuliers concrets. On peut vouloir s'interroger sur le statut ontologique d'êtres collectifs ou immatériels. Par exemple : « Quelle est l'essence du peuple français ? », « Quelle est l'essence de la *Neuvième Symphonie* de Beethoven ? », « Y a-t-il une loi morale ? ». La métaphysique se demande *qu'est-ce qui existe*, l'ontologie se demande *qu'est-ce qu'exister*. Bien sûr, les deux questions sont liées. Les questions ontologiques les plus générales pouvant être dites se rapporter à un être « suprême » identifié au réel en son entier (onto-théologie).

---

7 Ici on s'interroge sur le fait même d'être, distingué de la non-existence. On recherche le point commun à tout ce qui est, à tous les étants.

8 Ici on s'interroge donc plutôt sur ce qui constitue le « mobilier » du monde. On recherche de quoi est constitué le monde, de quels types d'êtres il est formé.

Si on veut, la métaphysique est à l'ontologie ce que les sciences sont à l'épistémologie. Comme les sciences, la métaphysique cherche à acquérir une connaissance sur le monde, mais alors que les sciences partent des faits situés dans l'espace et le temps, et en recherche les causes antécédentes dans l'espace et le temps, selon un processus d'abduction, de déduction, puis d'induction, la métaphysique, bien que partant aussi des faits situés dans l'espace et le temps, recherche, non pas leurs causes antécédentes, mais à les subsumer à un principe qui transcende l'espace et le temps. Pour cela, elle tente une inférence à la meilleure explication (qui est une forme d'*abduction créatrice*), mais sans la tester (sans *méta-abduction*), pour la simple raison que l'explication inférée est non testable. Un recours à l'analogie est nécessaire pour comprendre les énoncés métaphysiques, les poser et les sélectionner.

Le Mythe de la Caverne de Platon illustre bien ce raisonnement analogique. Les ombres sur le mur sont aux objets éclairés par le Soleil, ce que les objets éclairés par le Soleil sont aux Idées, lesquelles sont, dans la pensée platonicienne, le soubassement réel, abstrait et permanent des apparences sensibles changeantes. On ne peut donc véritablement connaître le réel en soi, indépendamment de notre capacité à le connaître. On peut, au mieux, l'imaginer, le ressentir, le supposer, le conjecturer, etc.

Souvent, la spéculation est utilisée. **La méthode spéculative** consiste à inverser notre rapport au réel (ou du moins à ce que l'on tiendra pour tel). Notre appréhension du réel est génératrice d'idées à son sujet. La méthode spéculative consiste à prendre nos idées, non pas pour un pâle reflet et résumé de la réalité, mais pour la réalité elle-même, dont chaque phénomène sera compris comme une application. Il faut mentionner aussi **la méthode dialectique** qui recherche la vérité métaphysique par la synthèse des opinions

contraires, ce qui permet de dépasser ceux-ci et d'atteindre une vérité sur laquelle tout le monde tombera d'accord. Concrètement, c'est le type de processus qui a lieu lors de débats ou de discussions entre des tenants d'options opposées. Prenons par exemple le problème de l'existence de Dieu. D'un côté, de soi-disant miracles, l'argument du dessein, la révélation, etc. suggèrent l'existence d'un Dieu. D'un autre côté, le problème de l'existence du mal suggère l'inexistence de Dieu. La contradiction pourra être résolue, par exemple, en supposant que Dieu existe bel et bien, mais que le mal a une utilité cachée : par exemple, il peut servir au perfectionnement des âmes, et que s'il est intolérable du point de vue humain, il ne touche guère l'âme. Ensuite, on pourra défendre l'idée suivant laquelle le mal ne nous affecte pas vraiment car nous sommes avant tout une âme, etc. A partir d'une contradiction apparente, la méthode dialectique va donc engendrer de nouvelles vérités métaphysiques. **La méthode transcendantale**, initiée par Emmanuel Kant, est une autre méthode métaphysique très en vogue, surtout au sein de l'apologétique chrétienne actuelle. Il s'agit de partir d'une certaine activité admise, puis de remonter à ses conditions de possibilité. Par exemple, supposons que l'on s'interroge sur l'existence ou non du langage. On pourrait faire remarquer que dire : « Le langage n'existe pas », suppose de faire usage du langage, et donc qu'il n'est pas logique de faire cette affirmation. Par conséquent, le langage existe nécessairement. Quelques autres méthodes utilisables en métaphysique sont :

- **La méthode historique :** Nos problèmes ont déjà été posés et des réponses ont déjà été apportées. Certaines réponses eurent peu de succès, ne se répandirent guère, furent vite abandonnées. D'autres, au contraire, influencèrent profondément l'humanité, furent même proposées indépendamment par plusieurs philosophes, et nourrissent encore nos réflexions aujourd'hui. Bien sûr, en soi, l'ancienneté et la popularité d'une

hypothèse philosophique n'est pas une preuve de sa validité, et un raisonnement correct sera toujours préférable. Néanmoins, on peut parier sur une certaine « sélection culturelle » des idées[9]. Celles qui survivent et se reproduisent (en mutant parfois quelque peu) sont celles qui sont les mieux adaptées à leur environnement, qui répondent au mieux aux besoins des Hommes, qui se rapprochent le plus de la vérité. De plus, au lieu de réinventer la roue à chaque fois, il peut être bon de se pencher sur le passé, qui est un prodigieux réservoir de matière à réflexion, de cadres logiques et conceptuels, de problématiques déjà explorées, de solutions déjà élaborées qui ne demandent parfois qu'une simple petite retouche. Examiner l'histoire de la philosophie permet aussi de dégager le sens qu'emprunte le cours de la pensée et vers où il a des chances de se diriger à l'avenir.

- **La méthode phénoméniste :** Elle cherche à faire une analyse du donné de l'expérience en lui-même, en mettant le monde extérieur entre parenthèses. *La Construction Logique du Monde*, par Rudolf Carnap[10], est un bon exemple d'ouvrage métaphysique (bien que l'auteur serait sans doute interloqué d'entendre dire qu'il fait de la métaphysique) qui exploite la méthode phénoméniste. Il part d'une analyse aprioriste des données de l'expérience subjective, qu'il décompose en sense data. Puis il essaye de traduire toute réalité du monde extérieur en combinaison de sense data descriptibles sous forme d'énoncés dits protocolaires.

---

9 Cf. la théorie des « mèmes » de Richard Dawkins : DAWKINS, Richard. *Le Gène Egoïste*.

10 CARNAP, Rudolf. La Construction Logique du Monde.

- **La méthode analytique :** Inspirée par la philosophie du second Wittgenstein, elle se base sur une étude du langage. Souvent, son approche des problèmes métaphysiques consiste à montrer… que ce ne sont pas de véritables problèmes, car ils naissent d'un usage incorrect du langage. Le rôle de la philosophie est thérapeutique. Elle doit guérir notre esprit des pseudo-problèmes métaphysiques insolubles qui nous tourmentent en montrant que ce ne sont pas de véritables problèmes mais des non-sens. Le paradigme est le suivant : le langage est ambigu (les mêmes mots peuvent vouloir dire des choses différentes selon les contextes), teinté de subjectivité (des expressions comme « moi », « ici », « maintenant », « hier », « à droite », etc. sont des indexicaux dont le sens varie avec le locuteur et la situation spatio-temporelle) et ses concepts sont vagues (par exemple, à partir de quelle taille est-on petit ?). La métaphysique tente néanmoins d'utiliser le langage ordinaire, comme s'il était clair et objectif, ou du moins s'invente un jargon propre tentant de fonder ou d'expliquer le langage ordinaire, mais sans sortir de celui-ci (par exemple, le *Dasein* sera l'être-là, *Dieu* sera l'être nécessaire, le *Néant* sera l'absence d'être, etc.). Comment le pourrait-elle, d'ailleurs ?[11] Ensuite, en appliquant une logique rigoureuse à ce qui ne l'est pas, on aboutit à des paradoxes, des problèmes, des apories, qui tourmentent les métaphysiciens. Par exemple, les problèmes sorites : Un

[11] En effet, il y a deux manières de définir un mot : soit à partir d'autres mots, soit en insérant ce mot dans une pratique (par exemple, montrer du doigt l'objet désigné par ce mot tout en le prononçant). On parle alors de définitions ostensives ou déictiques. Toutefois, la critique du langage privé nous montrera que la possibilité de ce second type de définitions requiert nécessairement un accès public à l'objet et à la cérémonie par laquelle on le nomme. Or, de nombreuses notions de la métaphysique ne présentent pas les conditions requises pour pouvoir donner lieu à des définitions ostensives, déictiques. Elles renvoient à des supposées réalités qui existeraient par-delà le domaine du publiquement observable.

homme de 2 m est grand. Si un homme de 2m est grand alors un homme de 1m99 est grand. Un homme de 1m99 est grand. Si un homme de 1m99 est grand, alors un homme de 1m98 est grand. Etc. Un homme de 1m est grand. Le raisonnement est d'une logique implacable, toutes les prémisses sont vraies, et pourtant la conclusion est fausse. Dans ce cas, c'est parce que la définition de « grand » est vague, ambiguë, relative, et que l'on fait comme si elle n'y était pas. Bien sûr, l'exemple n'est pas impressionnant, car personne ne prétend que le terme de « grand » ait un sens bien défini. Mais il arrive que l'on se fasse avoir par certains mots ou certaines expressions, qui nous paraissent précis. Certains de ces mots et expressions sont directement issus du langage ordinaire (« jeu », « chien », « moi », etc.), d'autres sont des mots et expressions techniques définis exclusivement à partir du langage ordinaire (« sense data », « ego transcendantal », « doute hyperbolique », « visée noématique », etc.) dans la fausse illusion qu'un assemblage logique de mots flous peut générer une définition précise. En métaphysique, la méthode analytique sera donc souvent regardée comme une entreprise de déconstruction. A travers elle, « *[n]ous* reconduisons les mots de leur usage métaphysique à leur usage quotidien. »[12]. Cependant, on voit naître, depuis quelques années, une métaphysique analytique proprement dite. Ce courant, au lieu de rejeter le langage ordinaire comme vague, ambigu et relatif, effectue une sorte de « renversement copernicien » : peut-être que notre langage est au contraire très précis, mais que c'est la réalité qui est vague. La métaphysique analytique part donc du langage ordinaire tel qu'il est, et en tire toutes les implications

[12] WITTGENSTEIN, Ludwig. *Recherches philosophiques*, § 116, p. 85.

métaphysiques et ontologiques. La métaphysique rationnelle à la Parménide, se trouve ainsi remplacée par une métaphysique héraclitienne du devenir, du changement, du vague, de l'incertain, plus proche de notre sens commun. Les prémisses de cette conception de la métaphysique se trouvent déjà chez Wittgenstein, lorsqu'il affirme que « [l]a philosophie ne doit en aucune manière porter atteinte à l'usage effectif du langage, elle ne peut donc, en fin de compte, que le **décrire**. Car elle ne peut pas non plus le fonder. **Elle laisse toutes choses en l'état.** »[13]. Plus loin, il écrit : « La philosophie se contente de placer toute chose devant nous, sans rien expliquer ni déduire. — Comme tout est là, offert à la vue, il n'y a rien à expliquer. Car ce qui est en quelque façon caché ne nous intéresse pas. »[14]. Cette fonction descriptive de la philosophie n'est en rien opposée à sa fonction thérapeutique. Dissiper les illusions linguistiques de la métaphysique contribue à éclaircir le véritable sens des concepts, à mieux décrire le langage ordinaire, débarrassé des équivoques et du jargon de la métaphysique classique. Peter Strawson, dans *Individuals*, parle de métaphysique *descriptive*, par opposition à une métaphysique qui serait *révisionniste* (la métaphysique classique, illustrée par les méthodes précédentes). La métaphysique révisionniste cherche à expliquer le monde, à lui conférer un sens transcendant. Elle ajoute, à l'ameublement du monde, à côté des entités ordinaires et naturelles – parfois à la place, ces dernières n'étant alors considérées que comme de simples phénomènes de surface – un ensemble d'entités « exotiques »

13 WITTGENSTEIN, Ludwig. *Recherches philosophiques*, § 124, p. 87. La mise en forme grasse a été ajoutée par nos soins.

14 WITTGENSTEIN, Ludwig. *Recherches philosophiques*, § 126, p. 88.

> supposées rendrent compte de l'existence du monde tel qu'on l'expérimente. Elle cherche à réviser nos concepts, à changer nos croyances générales sur le monde. En revanche, la métaphysique descriptive, comme son nom l'indique, se contente, de décrire le monde tel qu'elle le trouve, sans rien lui ôter ou lui rajouter.

Aucune question métaphysique ne peut recevoir une réponse définitive. Il y a une part nécessaire de préférence subjective et de recours à l'analogie, plus qu'à la logique dans le choix d'une option métaphysique particulière.

Les questions scientifiques peuvent, selon une certaine perspective qualifiée de réaliste, être vues aussi comme des questions métaphysiques. On peut admettre, selon le réalisme, que les théories scientifiques décrivent une réalité qui existe par elle-même, indépendamment de notre capacité à l'expérimenter et à la concevoir (et cela quand bien même nous l'expérimentons et la concevons en effet). Ce réalisme s'inscrit dans le cadre plus large d'une ontologie et métaphysique analytiques, qui prennent au sérieux ce que le langage nous révèle sur le monde. Ici, c'est le langage scientifique qui est concerné au premier chef.

Toutefois, d'autres ontologies et métaphysiques analytiques sont possibles. Ainsi, dans le cadre de notre étude, nous nous intéresserons particulièrement aux présupposés ontologiques impliqués par l'usage du concept de personne. Nous les ferons dialoguer avec ceux issus des diverses sciences, notamment des sciences cognitives qui, dans leur quête de l'esprit, peuvent avoir des choses intéressantes à nous dire concernant la personne ; et de la biologie, car la question de l'identité personnelle s'adresse aussi à elle, une personne

pouvant aussi être vue comme un individu biologique particulier, et pas seulement comme un pur esprit. Outre les sciences dites « dures », les sciences humaines et sociales (sciences « molles » ?) seront intéressantes à questionner, d'un point de vue ontologique et métaphysique, même si nous ne nous y attarderons pas aussi longtemps. Mais il est important de souligner le rôle non négligeable que peut présenter l'environnement spatial de l'individu et sa place dans un certain schéma relationnel dans la définition de son identité personnelle. Nous montrerons que les sciences de la complexité ouvrent de nouveaux chemins de compréhension de l'identité personnelle, notamment au regard des questionnements métaphysiques traditionnels sur le sens de la vie, l'origine et la destinée du « moi ».

Nous nous concentrerons autour de trois axes dialectiques :

- Un axe opposant les conceptions réductionnistes, voire éliminativistes, et antiréductionnistes de la personne

- Un axe opposant les conceptions psychologiques et biologiques de la personne

- Un axe opposant les conceptions internalistes et externalistes de la personne

## *B. La méthodologie des expériences de pensée*

La manière dont les recherches philosophiques sur la question de l'identité personnelle sont menées a de quoi surprendre. Depuis Locke déjà, et surtout depuis les 50

ou 60 dernières années, la littérature philosophique se concentre sur des arguments d'un type bien particulier : les expériences de pensée. Ces arguments se basent sur des scénarios imaginaires, dans lesquels on cherche à illustrer une séparation – que l'on tient pour physiquement, métaphysiquement et/ou conceptuellement possible – de caractéristiques du réel qui, dans tous ou presque tous les cas actuels coïncident. La présentation de ces cas fictifs est supposée évoquer un certain nombre d'intuitions chez le lecteur, au regard de la manière dont il conviendrait de les décrire ou d'y réagir. Ces intuitions sont supposées refléter une certaine vérité métaphysique, ou plus prosaïquement, nous permettre de mieux prendre conscience de nos croyances, et de la manière dont nous utilisons nos concepts.[15]

## 1. De la nature des expériences de pensée

Les expériences de pensée existent depuis l'Antiquité. Par exemple, dans *De la Nature des Choses*, Lucrèce[16] tente de montrer, par le biais d'une expérience de pensée, que l'espace est infini : S'il y avait une prétendue limite à l'Univers, nous pourrions lancer une lance vers elle. Si la lance la traverse, alors ce n'est pas une limite après tout, et si la lance rebondit, alors il doit y avoir quelque chose derrière la supposée bordure de l'espace, un mur cosmique qui stoppe la lance, un mur qui est lui-même dans l'espace. Quel que soit ce qui arriverait, il n'y a pas de bord à l'univers, l'espace est infini.

Cet exemple illustre plusieurs des caractéristiques communes aux expériences de pensée : nous visualisons une situation, nous menons une opération sur cette situation

---

15 Pour ce sous-chapitre, nous nous sommes largement inspiré de : BROWN, James Robert. "Thought Experiments". *In: Stanford Encyclopedia of Philosophy*.

16 LUCRECE, De la Nature des Choses.

visualisée, nous regardons ce qui arrive. Cela illustre aussi leur faillibilité. Ainsi, de nos jours, nous avons appris à conceptualiser l'espace de telle manière qu'il puisse être à la fois fini et illimité. Considérons par exemple un cercle, qui est un espace unidimensionnel : Tant que nous tournons autour, il n'y a pas de limite, mais il est néanmoins fini. L'univers pourrait être une version multidimensionnelle de ce qui se passe avec le cercle.

Souvent, une réelle expérience qui serait l'analogue d'une expérience de pensée serait impossible pour des raisons physiques, technologiques ou financières, mais cela n'est pas très important pour qu'une expérience de pensée puisse jouer son rôle.

En effet, il semble qu'à travers les expériences de pensée, nous serions en mesure d'atteindre une vérité concernant la nature, juste par la pensée, et c'est ce qui fait son grand intérêt pour la philosophie. Comment est-il possible d'apprendre apparemment de nouvelles choses concernant la nature sans de nouvelles données empiriques ?

## 2. Fonctionnement et épistémologie des expériences de pensée

Pour John Norton[17], toute expérience de pensée n'est réellement qu'un argument (possiblement déguisé). Elle commence avec des prémisses fondées dans l'expérience, suit des règles déductives ou inductives d'inférence et arrive à sa conclusion. Le caractère « imagé » de l'expérience de pensée qui lui confère son côté « expérimental » peut être psychologiquement facilitant (par exemple, il aide à mieux comprendre, à exemplifier, à retenir l'argument, à en percevoir la justesse ; il lui apporte aussi un côté ludique agréable,

[17] NORTON, John. "On Thought Experiments: Is There More to the Argument?". Et : NORTON, John. "Why Thought Experiments Do Not Transcend Empiricism". *In:* HITCHCOCK, Christopher (ed.). *Contemporary Debates in the Philosophy of Science.*

moins abrupt qu'une démonstration froide et sèche), mais il est strictement redondant. Donc, dans une expérience de pensée, pour Norton, nous n'allons jamais au-delà des prémisses empiriques d'une manière qui prêterait le flanc à des objections de la part de l'empirisme.

Pour Ernst Mach[18], les expériences de pensée mobilisent une « connaissance instinctive », non systématisée, implicite, acquise par l'expérience au cours de notre vie. Cette connaissance se révèle au contact de certaines situations qui la mobilise. A sa suite, Thomas Kuhn[19] et Tamar Szabò Gendler[20], voient l'expérience de pensée comme une sorte d'insight intellectuel, qui permet – sans l'aide de nouvelles données empiriques – de re-conceptualiser plus adéquatement l'Univers, de reconstruire l'espace du problème. Cette théorie n'est pas incompatible avec la précédente. Dans un contexte de découverte, une représentation imagée, un raisonnement analogique, une démarche synthétique et globale d'approche d'un problème, peuvent aider à trouver plus facilement sa solution qu'une description symbolique et abstraite, un raisonnement logique, une approche analytique et linéaire, et cela vaut également dans un contexte pédagogique, pour mieux faire comprendre une notion ou une théorie. Dans un contexte de justification, il pourrait néanmoins être possible, à chaque fois, de reconstruire l'expérience de pensée sous la forme d'un argument basé sur des suppositions explicites ou tacites, et la croyance en la conclusion de l'expérience de pensée pourrait n'être justifiée que dans la mesure où l'argument reconstruit pourrait justifier cette conclusion. De plus, dans un contexte de

---

18 MACH, Ernst. *The Science of Mechanics*.

19 KUHN, Thomas. "A Function for Thought Experiments". *In:* KUHN, Thomas (ed.). *The Essential Tension*, p. 240-265.

20 GENDLER, Tamar Szabò, "Galileo and the Indispensability of Scientific Thought Experiment".

découverte, il se pourrait que le déroulement d'une expérience de pensée puisse toujours se ramener au déploiement d'un argument logique standard. L'intérêt de l'expérience de pensée en tant que telle serait donc purement heuristique, pédagogique et psychologique.

Une troisième option épistémologique est toutefois en profond désaccord avec l'empirisme de John Norton. Il s'agit du platonisme de James Robert Brown[21]. Pour Brown, les lois de la nature résident dans un monde platonique, et que, des expériences de pensée d'un certain type, nous permettent de les appréhender directement. Ainsi, nous allons bien au-delà des données empiriques, et atteignons une connaissance *a priori* de la nature. Par exemple, Galilée a montré que tous les corps tombent à la même vitesse avec une brillante expérience de pensée qui commença par détruire le compte-rendu aristotélicien de la chute des corps. Selon ce dernier, les corps lourds (H) tombent plus vite que les corps légers (L) : H > L. Mais considérons un lourd boulet de canon (H) et une légère balle de fusil (L), attachés ensemble pour former un objet composite (H+L). Ce dernier doit tomber plus vite que le boulet de canon seul. Toutefois, l'objet composite devrait aussi tomber moins vite que le boulet de canon seul, puisque la partie légère agira comme un frein sur la partie lourde. Nous avons donc une contradiction : H+L > H et H > H+L. Cela met en pièces la théorie aristotélicienne. Mais il y a un plus, puisque le bon compte rendu est à présent évident : tous les objets tombent à la même vitesse : H = L = H+L. Cela peut être considéré comme une connaissance *a priori* (quoique encore faillible) de la nature, puisqu'il n'y a aucune nouvelle donnée empirique impliquée et que la conclusion ne dérive pas des anciennes données et n'est pas non plus une sorte de vérité logique.

---

21 BROWN, James Robert. "Peeking into Plato's Heaven".

## 3. Types d'expériences de pensée

Certaines expériences de pensée sont utilisées dans un but pédagogique, pour mieux faire comprendre un concept ou une théorie complexe. C'est par exemple le cas de l'expérience de pensée du démon de Laplace, conçue en vue de faire comprendre la thèse du déterminisme, en physique. On peut aussi utiliser les expériences de pensée comme pompes à intuitions (l'expression est de Dennett) en vue de résoudre certains problèmes. On raconte ainsi qu'Einstein eut l'idée de la théorie de la relativité, en s'imaginant assis sur un rayon de lumière. Dans ce dernier cas, l'expérience de pensée n'a pas valeur de validation, mais seulement d'aide à la recherche, de stimulant intellectuel si on veut. Enfin, on peut voir toute la littérature de science-fiction comme un gigantesque réservoir d'expériences de pensée. A partir de l'état actuel de la science, on projette, par extrapolation, ce qu'il pourrait en être dans plusieurs décennies. Toutes ces expériences de pensée présentent comme point commun qu'elles n'ont pas la prétention de constituer des validations ou invalidations d'une théorie. En cela, elles ne sont pas très problématiques.

Toutefois, il existe un autre genre d'expérience de pensée, qui ressemble analogiquement aux expériences menées en sciences en cela qu'elles sont censées permettre de vérifier – en un certain sens – l'adéquation d'une théorie à la réalité.

D'après James Brown, il existe deux grands types de ce genre d'expériences de pensée, qui se subdivisent chacun en plusieurs sous-types :

1. Expériences de pensée négatives, ou destructives : Elles sont utilisés négativement pour miner une théorie. Cela peut se faire de diverses manières :

a. Certaines expériences de pensée cherchent à montrer une contradiction dans une théorie, et donc la réfuter. C'est par exemple le cas avec l'expérience de pensée de Lucrèce contre l'idée d'un Univers fini. Les deux possibilités : la lance traverse la limite, la lance rebondit, sont toutes deux incompatibles avec l'idée d'un Univers fini et est supposé démontrer le caractère contradictoire de cette idée.

b. D'autres expériences de pensée cherchent à montrer que la théorie en question est en conflit avec d'autres de nos croyances. C'est souvent le cas des expériences de pensée concernant l'identité personnelle. Les croyances sont ici les croyances du sens commun. On choisira alors, ou bien de privilégier ces croyances, et de rejeter la théorie de l'identité personnelle testée, ou bien de rejeter nos croyances du sens commun (ou du moins de restreindre leur validité aux cas « normaux ») et de garder la théorie, ou bien de rejeter les deux et de suggérer une re-conceptualisation de l'ensemble qui permette de régler le conflit.

c. Les contre expériences de pensée cherchent à miner une supposition centrale ou une prémisse d'une expérience de pensée. Il s'agit d'argumenter contre une expérience de pensée à l'aide d'une autre expérience de pensée, supposément mieux conçue. On pourra consulter plusieurs exemples de cette stratégie (concernant le problème des qualia) dans l'article de Daniel Dennett, « Quining Qualia »[22].

---

22 DENNETT, Daniel Clement. "Quining Qualia". *In:* LYCAN, William G. *Mind and Cognition*, p. 519-548.

2. Les expériences de pensée positives, ou constructives : Elles sont utilisées positivement pour établir, appuyer ou illustrer une théorie, ou encore pour mettre en exergue un phénomène à expliquer. Il existe plusieurs sous-types :

    a. Les expériences de pensée médiatrices : Ce sont celles qui facilitent l'établissement d'une conclusion, à partir d'une théorie bien articulée spécifique. Elles fonctionnent en fournissant une sorte d'illustration qui rend les propositions d'une théorie claires et évidentes. Dans de tels cas, les expériences de pensée sont une sorte d'aide heuristique. Un résultat peut être déjà bien établi, mais l'expérience de pensée en donnera une compréhension intuitive.

    b. Les expériences de pensée conjecturelles : Ce sont celles pour lesquelles réfléchir à un scénario imaginaire va faire en sorte que nous prêtions attention à un phénomène pour lequel nous fournirons alors une explication théorique. Par exemple, l'expérience de pensée de Frank Jackson avec la scientifique Mary, confinée depuis sa naissance dans un univers entièrement en noir et blanc, est supposée nous faire prendre conscience de l'existence des *qualia*.

    c. Les expériences de pensée directes : Ce sont celles qui fournissent directement une théorie bien établie.

Notons que James Brown appelle « expériences de pensée platoniques », des expériences de pensée qui sont à la fois destructives et constructives directes, car, selon lui, elles apporteraient une connaissance *a priori* de la nature.

# III. Eléments de géographie conceptuelle

## *A. Espace du problème*

### 1. Identité personnelle versus identité individuelle

Les concepts de personne et d'individu sont souvent confondus dans la langue courante et utilisés indifféremment pour désigner soit ce qu'il convient d'appeler véritablement une personne, soit pour ce qu'il convient d'appeler seulement un individu.

Les philosophes eux-mêmes ne sont pas toujours conscients du sens réel de ces mots, et beaucoup de controverses philosophiques en viennent à n'être que des problèmes de définitions. Si notre langue avait un mot précis pour chaque chose et que chaque mot ne référait qu'à une chose unique, sans doute que de nombreux débats n'auraient pas lieu d'être et que des philosophes pensant défendre des vues diamétralement opposées se rendraient compte avec étonnement qu'ils tombent d'accord…

Pour notre part, nous nous en tiendrons au sens habituel et étymologique des termes d'individus et de personnes. Un individu (lat. *individuum*, qui ne peut pas être divisé), est un être qui ne peut être divisé sans perdre par là même les caractéristiques qui lui sont propres. C'est donc quelque chose de très concret, pas forcément quelque chose de pensant mais au moins un être vivant. Un homme est un individu, un chien en est un autre, ainsi qu'un

arbre ou une paramécie. Boèce, ainsi que Platon, Aristote et leurs disciples parlaient plutôt d'*hypostase*. Le terme est un peu plus connoté métaphysiquement. Une hypostase, c'est une substance ou une nature, individuée.

Une personne (lat. *persona*, désignait, chez Boèce[23], le masque porté par les acteurs lorsqu'ils jouaient leur rôle au théâtre) désignera le rôle que nous jouons aux autres et à nous-mêmes, sans systématiquement en prendre conscience. Ce n'est donc pas une nature ou une essence, mais relève plutôt d'un acquis qui peut varier selon les individus. Ce n'est pas un état, c'est une disposition. Notre personne, c'est comment nous nous définissons, l'ensemble plus ou moins stable des pensées en première personne que nous avons de nous-même. Etre une personne implique bien sûr que nous soyons donc déjà des êtres conscients. Mais il ne suffit pas d'être conscient de son environnement pour être une personne (sinon un chien ou même une poule seraient des personnes). Il faut être capable d'un certain type de pensée, que Stéphane Chauvier[24] appelle pensées égologiques. Ce sont les pensées dont **je** suis le sujet et l'objet. Par exemple : « Je suis un grand timide », « Je préfère la vanille à la pistache », « J'ai toujours cru que tu n'aimais que moi », sont des pensées égologiques. Enfin, pour être une personne, il faut des préférences, des motivations, etc. ce qui nous permet d'agir selon des buts, en fonction de motifs, de normes, de valeurs, etc. que nous nous fixons ou que l'on nous impose. Un être qui subirait passivement son environnement sans faire preuve de spontanéité ne serait pas vraiment une personne… On pourrait lui faire subir ce que l'on voudrait sans jamais lui nuire, puisqu'il ne désirerait pas autre chose que la situation telle qu'elle est, quelle qu'elle soit…

---

23 BOECE. *Courts traités de théologie*.

24 CHAUVIER, Stéphane. *Qu'est-ce qu'une personne ?*

Ayant défini ce qu'est un individu et ce qu'est une personne, une première série de questions qui se posent est la suivante : Toute personne est-elle forcément un individu ? Existe-t-il des personnes qui ne sont pas humaines ou des humains qui ne sont pas des personnes ? Quand un être humain devient-il une personne et quand cesse-t-il d'en être une ? L'identité personnelle présuppose-t-elle l'identité individuelle ? Se ramène-t-elle à l'identité individuelle ? Peut-on être la même personne tout en changeant radicalement de personnalité ? Un même individu peut-il être plusieurs personnes à la fois ou successivement ? Une même personne peut-elle exister dans plusieurs individus à la fois ou successivement ?

## 2. Trois axes au problème

On peut, pour se représenter de manière imagée les choses, voir les personnes comme des polygones en trois dimensions évoluant à travers un espace vectoriel à trois dimensions : le temps, l'espace et la modalité. Dans certains cas, on a les coordonnées de chaque point de ce polygone, et on cherche la fonction qui les relie. Dans d'autres cas, on a une fonction, et on cherche à savoir si elle décrit bien l'emplacement des points connus du polygone, et ce qu'elle nous permet d'interpoler et d'extrapoler en ce qui concerne d'autres points hypothétiques du polygone lorsque l'espace, le temps et la modalité prennent d'autres valeurs. La fonction, c'est l'identité personnelle, ou du moins un certain critère ou ensemble de critères dont on suppose et espère qu'ils coïncideront avec ce qu'il convient d'entendre par *identité personnelle*. Voyons un peu plus précisément, pour chaque dimension, quels sont les problèmes spécifiques qui se posent.

## a) Temps

Traditionnellement, et c'est encore souvent le cas dans la philosophie analytique anglo-saxonne, le problème de l'identité personnelle est assimilé au seul problème de l'identité personnelle *à travers le temps*. Quelles sont les conditions nécessaires et suffisantes pour qu'une personne A vivant à un temps t1 soit la même (au sens de l'identité numérique, donc au sens de : une seule et même) personne qu'une personne B vivant à un temps t2 ? Il faut dire que le problème de l'identité personnelle s'est tout d'abord posé dans une perspective eschatologique. Locke[25], par exemple, se demandait si, le jour du Jugement Dernier, Dieu pouvait condamner à l'Enfer un individu n'ayant plus aucun souvenir des crimes qu'il aurait commis. La préservation de la mémoire est-elle une condition nécessaire et (éventuellement) suffisante à l'identité d'une personne à travers le temps ? Y a-t-il d'autres critères que l'on peut faire intervenir ? C'est ce que nous verrons plus tard.

## b) Espace

Ce que l'on pourrait appeler *identité personnelle à travers l'espace* doit être compris de la manière suivante (et non pas au sens d'une identité numérique, mais plutôt au sens méréologique d'une relation d'appartenance d'une partie à un tout) : soit une personne P, soit une fraction d'espace (x ; y ; z), quelles sont les conditions nécessaires et suffisantes pour que la fraction d'espace (x ; y ; z) fasse partie de l'hypostase qui individue P ? La question peut sembler triviale : les individus ne sont pas des nuages, se sont des entités bien séparées les unes des autres, et à chacun correspond au plus une personne. Toutefois, nous verrons des cas (réels ou fictifs) où il semble qu'une personne existe simultanément en

[25] LOCKE, John. *Essai sur l'entendement humain*. Livre II, Chap. 27, § 22, p. 535-536.

plusieurs individus, ou bien où plusieurs personnes semblent cohabiter simultanément au sein d'un seul individu (des cas réels de ce genre sont représentés par les frères siamois et les patients split-brain).

### c) Modalité

Jusqu'à quel point une personne aurait-elle pu être différente de ce qu'elle est vraiment, tout en étant la même personne ? L'étude de l'identité personnelle à travers les mondes possibles cherche à répondre à cette question. A moins de considérer qu'il n'existe pas de faits contingents, c'est-à-dire que tout ce qui arrive devait arriver et que rien de ce qui n'arrive pas n'aurait pu arriver, cette question est pertinente. Aurions-nous pu, tout en préservant notre essence, naître 2 ans avant ou 3 mois après notre naissance ? Avoir eu d'autres parents ? Faire d'autres choix de vie ? Avoir été noir, indien, asiatique, etc. ? Aurions-nous pu être un chien, un artichaut, un caillou, une note de musique, un concept, etc. ?

## 3. Pourquoi est-ce bien moi qui suis moi ?

Il semble que la considération, ensemble, de l'axe spatial et modal du problème de l'identité personnelle puisse conduire à reconsidérer à nouveau un paradoxe qui a tourmenté et tourmente encore beaucoup de jeunes enfants en pleine période des questions et des « pourquoi ? », mais qui semble ne plus faire beaucoup d'impression sur les adultes que nous sommes, et qu'illustre ce passage de « La Chanson de l'Enfance » de l'ange Cassiel dans le film de Wim Wenders et Peter Handke, *les Ailes du Désir* :

« Quand l'enfant était un enfant,
C'était le temps des questions :
Pourquoi suis-je moi et pas toi ?
Pourquoi suis-je ici et pas là-bas ? »[26]

Parmi tous les êtres existants *maintenant*, comment se fait-il que je sois précisément celui que je suis à l'exclusion de tous les autres ? Aurais-je pu en être un autre ? Aurais-je pu être vous ? Jacques Chirac ? Albert Einstein ? Emmanuel Kant ? Le Pape ? Lady Di ? Le chien de la voisine ? La voiture de mon père ? La planète Mars ?

Il y a un caractère remarquable à ce problème en cela qu'il n'implique aucunement des mondes possibles où les événements *objectifs* se seraient déroulés autrement. En fait, deux mondes possibles pourraient, peut-être, être parfaitement indiscernables et ne différer qu'en cela seul que dans un des deux je serais, par exemple, Jacques Chirac, tandis que Jacques Chirac serait moi. De plus, elle n'implique même pas la notion de mondes possibles. Ainsi, on pourrait peut-être imaginer que, par miracle, je me retrouve à être Jacques Chirac les jours pairs et Mikaël Mugneret les jours impairs et qu'il en soit inversement pour Jacques Chirac, sans que cela soit le moins du monde décelable par quiconque, Jacques Chirac et moi y compris ! Stéphane Ferret, dans son ouvrage, *Le Philosophe et son Scalpel*[27], évite soigneusement cette question, tout en reconnaissant sa légitimité et son importance, car « ce dont on ne peut parler, il faut le taire ». Nous aimerions pourtant en dire deux mots, car contrairement à Stéphane Ferret, nous pensons que le problème peut faire l'objet d'un discours rationnel, mais qu'il est peut-être tout simplement mal posé. La plupart des définitions de la personne tentent d'appréhender ce

---

26 WENDERS, Wim et HANDKE, Peter. *Les Ailes du Désir.* [Script cinématographique]

27 FERRET, Stéphane. *Le Philosophe et son Scalpel*, Chap. 1 : « Le labyrinthe de l'identité personnelle », p. 20

concept d'une manière objective. Si la personne est un concept objectif, alors il y a toutes les chances pour qu'une personne en particulier soit un être objectif… même si elle est une pure subjectivité ! Mais rien ne nous garantit qu'une pure subjectivité puisse être saisie de manière objective, fut-ce par la seule pensée. C'est pourtant le présupposé constant de la plupart des réflexions sur l'identité personnelle, à quelques exceptions près, tel David Banach pour qui :

> « Une conception objective du monde, complète et précise, omet quelque chose d'important au sujet de mon identité : laquelle des nombreuses personnes dans ce tableau objectif est **moi**. »[28].

## 4. Problèmes de l'unité expérientielle et consciente

Descartes pensait que l'âme était indivisible. C'est d'ailleurs un des faits qui arguait pour son immatérialité et son immortalité. En effet, la matière est composée de parties, est donc divisible. Puisque l'âme est simple et par là même non divisible, il s'ensuit logiquement qu'elle n'est pas matérielle. Il est ensuite aisé de passer de cette conclusion à la suivante : l'âme est immortelle. Rappelons l'argument[29] :

1. Une chose ne peut cesser d'exister que par anéantissement ou par séparation de ses parties

2. L'âme est simple, sans parties

---

[28] BANACH, David. "Who do you think you are? Relations, Subjectivity, and the Identity of Persons".

[29] Cf. par exemple, BOULENGER, Auguste. *La doctrine catholique*, Tome 1 : « Le Dogme », 61, VIII : « Immortalité de l'âme humaine. », 2, a.

3. L'âme ne peut donc cesser d'exister par séparation de ses parties

4. L'observation de la nature ne nous donne pas d'exemples d'anéantissements et nous conduit donc à supposer que ceux-ci n'ont pas lieu dans l'ordre des choses

5. L'âme ne peut donc cesser d'exister par anéantissement

6. L'âme ne peut cesser d'exister, elle est immortelle

Mais pourquoi donc supposer que l'âme est une substance simple, indivisible ? A cause de l'unité de la conscience.

## a) Unité expérientielle et consciente

### *(1) Unité de la conscience*

Mais qu'est-ce que l'unité de la conscience ? Laissons la parole au neurobiologiste Axel Cleeremans :

> « Que veut-on dire exactement par "unité de la conscience" ? En première analyse, simplement le fait que notre conscience du monde présente, à chaque instant, un caractère intégré. Par exemple, alors que j'écris ces lignes, je suis conscient non seulement du contenu des pensées que j'exprime en formulant les phrases que j'écris, mais j'ai également et simultanément diverses autres expériences : Une expérience visuelle des couleurs qui forment le fond de mon écran, une expérience auditive du cliquetis rythmé que produisent les touches de mon clavier alors que je les enfonce successivement, une deuxième expérience auditive du bruit

> que fait le système de conditionnement d'air dans mon bureau, l'expérience d'une vague douleur dans ma cheville droite, et ainsi de suite. Chacune de ces expériences est distincte, mais chacune est également profondément intégrée — unifiée — avec les autres. A chaque instant, l'expérience que j'ai du monde forme donc une unité. Toutes les expériences qui nous sommes [sic] données sont vécues comme appartenant à un seul et même sujet. En outre, toutes les expériences que nous avons vécues au fil de notre existence sont également reconnues comme étant nos expériences. »[30]

Ainsi, l'unité de la conscience me place en observateur de mes propres expériences conscientes, comme si celles-ci m'étaient projetées sur un écran, un film. C'est ce que Dennett appelle « le théâtre cartésien ». Je peux explorer mon univers mental, mon monde, comme si j'en étais en quelque sorte séparé et extérieur, comme si j'étais une sorte d'homoncule doté d'un sens interne. A travers l'unité de la conscience, je me sens propriétaire de mes propres expériences conscientes.

Ce n'est qu'une image approximative. Il serait peut-être plus juste de dire que toutes mes expériences conscientes se présentent comme partie d'une seule expérience plus vaste[31]. Je n'ai pas seulement des sensations, j'ai aussi la sensation de ces sensations et de leurs rapports, les unes avec les autres. D'une certaine manière, cette sensation de mes sensations est une condition nécessaire à ce que j'ai ces mêmes sensations. Sinon, je n'en

---

30 CLEEREMANS, Axel. « L'Unité de la conscience ». *In :* CAZENAVE, Michel (Ed.). *De la science à la philosophie : Y a-t-il une unité de la connaissance ?* p. 149

31 La lecture de RAYMONT, Paul. "The Unity of Consciousness". *In: Stanford Encyclopedia of Philosophy*, nous fait prendre conscience du caractère ontologiquement orienté d'une telle formulation, puisqu'il y a deux conceptualisations possibles de l'unité de la conscience et de l'expérience : une considérant que la conscience est constituée de parties, une autre qui nie cela. Nous pensons néanmoins que cette formulation peut être conservée, au moins à titre purement *descriptif*, et en précisant que nous ne cherchons pas, pour notre part, en employant une telle formulation, à manifester un engagement ontologique pour l'une ou l'autre version de l'unité de la conscience et de l'expérience.

aurais qu'une seule. Cela ne veut pas dire que les autres n'existeraient pas, mais elles existeraient à travers une autre conscience qui ne serait pas **ma** conscience, elles ne seraient pas miennes. Elles n'existeraient pas *pour moi*, mais éventuellement *pour autrui*. Elles me concerneraient autant que les expériences conscientes d'autrui, ce qu'elles seraient, à vrai dire.

Mais l'unité de la conscience assure que, mon expérience, par exemple, d'un cor au pied, par rapport à mon expérience d'une chaleur dans le bras, n'est pas ce que l'expérience qualitativement identique du cor au pied de quelqu'un d'autre est par rapport à cette même expérience de la chaleur.

On peut tenter de formaliser ces intuitions concernant l'unité de la conscience de la manière suivante : on dira que deux expériences conscientes A et B sont unifiées si et seulement si l'expérience de A et l'expérience de B consiste en l'expérience de A **et** B.

L'unification de la conscience permet, sans sortir de l'immédiateté du donné, d'opérer à des comparaisons entre sensations, d'organiser donc, notre champ de conscience.[32]

### *(2) Unité du champ d'expérience*

Il convient d'être vigilant à la polysémie du mot « conscience », qui peut être assimilé : d'une part, au sentiment brut et irréfléchi d'existence (le fameux « *what it is like to be* » de Thomas Nagel[33]), autrement dit, à ce que nombre de philosophes qualifient de conscience phénoménale, expérientielle, qualitative, immédiate, subjective, etc. d'expérience vécue, phénoménale, qualitative, consciente, subjective, etc. de présence au monde, d'éveil, etc. et qui renvoie au problème des *qualia* ; et, d'autre part, à la conscience réfléchie, intentionnelle, attentionnelle, d'accès, de *monitoring*, à l'intention mais aussi l'attention, le sentiment du sentiment, l'expérience de l'expérience, la conscience de la conscience, le sens interne, la perception interne, « l'œil de l'esprit », etc. Pour marquer ces distinctions nous

---

32 On parle également d'unité de la conscience à propos des objets de la perception. En effet, une chaise, une table, un ordinateur, etc. sont appréhendés, par la conscience, comme des touts unifiés (structurellement, dynamiquement, et fonctionnellement), persistants, et organisés en parties, non pas comme des paquets de sensations fugaces. Les objets physiques ne sont des objets que pour autant qu'ils sont perçus comme des objets. Les figures ambiguës, comme celle du canard-lapin, ou encore le phénomène de paréidolie, illustrent bien ce point : que la perception des objets n'est pas qu'une imprégnation passive, par l'esprit, d'une réalité préexistante, mais une co-construction qui repose sur des interactions sensori-motrices entre le sujet cognitif et son environnement proximal. Le problème de Molyneux, qui se demandait si un aveugle-né à qui la vue serait rendue pourrait reconnaître visuellement un cube qu'il n'aurait, jusqu'à lors, appréhendé que tactilement, questionne un autre type d'unité de la conscience des objets : l'unité entre les représentations, selon diverses modalités sensorielles, d'un même objet ou type d'objet. Bien que non dénué d'intérêt, nous n'aborderons guère ce thème de l'unité des objets de la perception pour lui-même, puisque c'est bien plutôt le problème de l'unité du *sujet* de la perception qui nous préoccupe dans cette thèse.

33 Cf. NAGEL, Thomas, "What it is like to be a Bat?".

parlerons de **champ d'expérience**, vs de **conscience** proprement dite, respectivement. Ces distinctions ne doivent sans doute pas être radicalisées. Ainsi, pour Damasio, la conscience repose probablement sur l'expérience, dont elle n'est qu'une modification et sophistication.

Bien que l'on parle souvent d'unité à propos de la conscience seule, il semble bien que l'on puisse déjà parler d'une sorte d'unité, au stade des sensations inconscientes de l'expérience. Ainsi, même si je ne m'aperçois pas continuellement du sol sous mes pieds ou du fait que je respire, peut-on sérieusement douter que ces deux expériences sont reliées ? Ce n'est pas quelqu'un d'autre qui sent le sol sous mes pieds tandis que je ressentirais ma respiration. Sinon cette expérience du sol sous mes pieds pourrait très bien ne pas exister du tout que cela ne ferait aucune différence pour moi.

Il nous faut donc raisonnablement supposer que les sensations inconscientes, même si on ne les aperçoit pas, sont déjà unifiées. Là où l'unité de la conscience renvoie à l'unité, apparemment nécessaire de la personne, l'unité de l'expérience renvoie à l'unité, apparement nécessaire, de l'individu[34].

On pourrait songer à un contre-argument à cette thèse, à travers l'expérience, somme toute banale, de penser à une chose (ex. : composer un certain numéro de téléphone) et d'en faire une autre (ex. : composer un autre numéro de téléphone). Mais le caractère conflictuel de cette expérience présuppose nécessairement son unité, même si cette unité est paradoxale. En effet, supposons que Paul, habitant à Lille, pense à composer un certain

[34] Quelques auteurs, comme Parfit, remettent en partie question cette association. D'après eux, une même personne peut être le lieu de phénomènes conscients et expérientiels entièrement dissociés, en existant à travers deux individus. Cf. PARFIT, Derek, "Reasons and Persons".

numéro de téléphone, pendant que Pierre, habitant à Marseille, compose un autre numéro de téléphone, irions-nous comparer cette situation à la nôtre, lorsque nous pensons à un numéro de téléphone, et que, simultanément, nous composons un autre numéro de téléphone ? Cette situation, avec Paul et Pierre génère-t-elle, quelque part, une expérience d'étrangeté ? Paraît-elle conflictuelle ? Il ne peut y avoir conflit d'expériences que là où elles sont présentes simultanément. Quand bien même nous n'avons pas conscience de l'incohérence entre penser à un certain numéro de téléphone et en composer un autre, nous ressentons ces deux expériences simultanément, de manière intégrée.

## b) Unité synchronique et diachronique

L'unité de la conscience et de l'expérience se déploie selon au moins deux axes. Un axe spatial et un axe temporel. On peut ainsi parler d'unité synchronique et diachronique, respectivement. L'existence de cette unité synchronique est la condition même pour que nous puissions parler de spatialité et de sentiment de l'espace. L'existence de l'unité diachronique, quant à elle, est la condition nécessaire au sentiment intime de la durée, et donc à l'intelligibilité de la notion de temps.

### *(1) Unité synchronique*

Si l'on peut situer les objets de notre champ de conscience les uns par rapport aux autres, dire que l'un est à droite, à gauche, en bas ou en haut d'un autre, voire évaluer les distances qui les séparent, si l'on peut aussi dire d'un objet que l'on perçoit qu'il est plus gros ou plus petit qu'un autre, c'est parce que notre conscience est unifiée selon une première dimension : une dimension spatiale. Cette unité spatiale est également responsable

de l'identification entre nos différentes modalités sensorielles, comme par exemple la projection, dans notre représentation visuelle de notre corps, des sensations corporelles (douleur, plaisir, froid, chaud, etc.). Cette identification n'a cependant rien d'absolu et peut être mise à mal par des expériences simples. En voici une : installez-vous sur une chaise, les yeux bandés. Pendant qu'avec votre bras gauche, vous essayerez de toucher, lorsque vous le souhaiterez, le nez d'un ami assis sur une chaise à votre gauche, un troisième larron, devra toucher votre nez, en veillant à faire correspondre ses touchers avec les vôtres. Au bout d'un certain temps, normalement, il est possible que vous ayez la sensation que votre nez a grandi ou s'est déplacé jusqu'au nez de votre voisin de gauche ![35]

En un autre sens, on pourra se demander pourquoi toutes les consciences de tous les individus ne sont pas unifiées en une seule grande conscience. D'un point de vue physicaliste, il semble évident que pour que cela soit possible, il faudrait que tous les individus soient physiquement reliés les uns aux autres. Même si nos cerveaux sont séparés, ainsi, en général, que nos corps (si l'on excepte le cas des frères siamois), on ne peut pourtant pas dire que les individus sont des systèmes isolés. Des liens plus ou moins lâches existent entre eux. Peut-être ne sont-ils pas suffisants ? Mais est-ce toujours le cas ? On peut en douter : nous vivons dans une ère de la communication intensive et à très grande vitesse, notamment grâce au développement de l'Internet. Alors, serait-il possible que finalement une telle conscience existe ?...

### *(2) Unité diachronique*

[35] Cf. RAMACHANDRAN, Vilayanur S. et BLAKESLEE, Sandra. *Le Fantôme Intérieur.* Chapitre 3 : « Chasser le fantôme », p. 87-91.

L'unité de la conscience existe également, même si cela n'est pas toujours remarqué, à travers le temps. Le présent est souvent représenté, en physique ou en mathématique, comme un point sur une droite, mais cela ne correspond manifestement pas à l'expérience vécue. Nous n'expérimentons pas un présent ponctuel mais un présent qui dure[36], un présent qui, paradoxalement, s'étend vers le passé (ce que Husserl[37] nomme *rétention*) et même vers le futur (ce que Husserl nomme *protention*), à travers le phénomène de l'attente, qui nous fait prendre conscience, par avance, de ce qui va probablement se passer, compte tenu de ce qui s'est déjà passé. Bien sûr, parfois les événements contrecarrent nos attentes, et nous sommes surpris, agréablement ou désagréablement…

L'unité diachronique de la conscience et de l'expérience est ce qui nous permet de prendre conscience, par exemple, d'une mélodie que l'on écoute, comme d'un tout indivisible. Nous avons une sensation de succession et non pas seulement une succession de sensations. C'est grâce à elle aussi que l'on perçoit le mouvement.

Peut-on vraiment parler d'unité diachronique de la conscience ? Si l'existence d'une unité de la conscience a toujours eu ses partisans et ses détracteurs, c'est en ce qui concerne la dimension temporelle supposée de cette unité que l'on rencontre le plus de scepticisme. L'apparence de cette unité ne peut-elle pas reposer simplement sur le souvenir rappelé

---

[36] D'un autre côté, cette unité locale (elle concerne le présent) doit être contrasté avec la disunité qui caractérise notre histoire globale. Pourquoi notre passé lointain ne peut être l'objet que d'un souvenir, et non d'une expérience directe, unifiée avec notre expérience présente, et pourquoi pas notre expérience future, en un présent couvrant toute notre vie ? Il semble que si cela devait arriver, ce serait une situation très curieuse : comment concevoir une action (semble-t-il nécessairement inscrite dans un temps bien précis) chez un être n'expérimentant pas le passage du temps ? Aurait-il les mêmes concepts de durée et de temps que nous ? Expérimenterait-il la patience, l'impatience, l'ennui, etc. ?

[37] Cf. HUSSERL, Edmund. *Leçons pour une phénoménologie de la conscience intime du temps*.

d'une forme de mémoire à très court terme ? Bien sûr, à travers le souvenir, nous avons un certain nombre d' « images » (au sens large, pas seulement des images visuelles) qui nous viennent à l'esprit avec l'impression que ces « images » correspondent à des événements que nous avons vécus dans le passé. Mais ces « images », si elles correspondent à des événements du passé, sont néanmoins vécues au présent. Il ne peut en être autrement. En effet, lorsque je dis : « Je me souviens de nos premiers baisers », cela veut dire que c'est *maintenant* que j'ai ce souvenir. Si je dis : « Je me souvenais de nos premiers baisers », alors il s'agit cette fois-ci, au mieux, de l'évocation d'un souvenir *actuel* ayant pour objet un autre souvenir situé, lui, dans le passé. De plus, l'évocation de souvenirs implique la mémoire. Or, lorsque l'on se demande si la conscience est diachroniquement unifiée, on se demande en fait si deux événements conscients temporellement disjoints peuvent être intégrés au sein d'une même représentation consciente vécue au présent. Autrement dit, on doit se poser la question suivante : le présent vécu a-t-il une épaisseur ou n'est-il qu'un instant ? Tout porte à croire, selon Kaplan[38], que le présent vécu a une épaisseur. En effet :

> « Prenons l'exemple de la sensation auditive : lorsque j'entends le dernier mot d'une phrase ou la dernière note d'une mélodie, j'ai présent à l'esprit les notes ou les mots précédents, sinon je n'aurais pas entendu une phrase ou une mélodie, mais une note ou un mot, ou même seulement une partie d'un mot ; et il y aurait eu succession de sensations et non sensation d'une succession. Mais cela signifie que les notes, les mots ou les sons du début de la durée, parce qu'ils sont précédents, sont passés et sont en même temps présents. De même, quand je mesure une durée, quand je dis que telle durée est plus longue que telle autre, c'est que j'ai présent à l'esprit, au moment où finit cette durée, tout ce qui précède cette fin, et cela signifie que tout ce qui précède cette fin, parce qu'il la précède, est passé et en même temps présent. Cela signifie, par conséquent, que le passé est présent. »

38 KAPLAN, Francis. « L'irréalité de la mort », *In* : LENOIR, Frédéric et DE TONNAC, Jean-Philippe, (Eds.). *La Mort et l'Immortalité. Encyclopédie des Savoirs et des Croyances*, p. 1434-1435.

Peut-on rendre compte de cette contradiction apparente en ayant recours à la mémoire ? C'était l'interprétation de Saint Augustin, qui avait déjà mis en évidence ce phénomène. Toutefois, Kaplan nous donne de bonnes raisons d'en douter :

> « Sans doute, le souvenir, au moment où on en prend conscience, est une réalité présente […]. Il n'y a donc rien d'étonnant à ce que toutes les notes passées de la mélodie ou tous les mots passés de la phrase que j'entends soient présents au moment où j'entends la dernière note de la mélodie ou le dernier mot de la phrase. Mais, ou je prends conscience successivement de chacune des réalités présentes correspondant à chacune des notes passées de la mélodie, à chacun des mots passés de la phrase, et nous nous retrouvons devant le même problème : ces prises de conscience ne sont pas présentes en même temps, il faut donc une nouvelle mémoire pour me rappeler, au moment de la prise de conscience du souvenir de la dernière note passée ou du dernier mot passé, des prises de conscience des souvenirs des notes ou des mots précédents – en quelque sorte des souvenirs de souvenirs et évidemment des souvenirs de souvenirs de souvenirs, et cela indéfiniment ; ou je [39]prends conscience en même temps, au même instant, de chacune des réalités présentes correspondant à chacune des notes passées et elles se télescoperaient, se confondraient : ce ne serait plus une phrase, mais une cacophonie ; ce ne serait plus une mélodie, mais, dans la meilleure des hypothèses, ce que l'on appelle un accord musical. Rétorquera-t-on qu'on peut les entendre en même temps et cependant les unes à part des autres comme, lorsqu'on voit un paysage en un instant et que pourtant on voit à part telle maison, tel arbre, telle colline, telle route, tel personnage ? Admettons qu'on puisse voir en un instant un paysage, qu'il ne faille pas du temps pour le balayer du regard – ce qui est peut-être contestable. Ce qui est sûr, du moins, c'est que l'ordre dans lequel on le balaie importe peu : on peut commencer par la colline de droite ou la maison de gauche. En revanche, pour une phrase ou une mélodie, l'ordre de succession est essentiel : « Pierre tue Jacques » et « Jacques tue Pierre » ont exactement les mêmes mots, mais dans un ordre de succession différent ; les notes d'une mélodie ne donnent pas la même

[39] KAPLAN, Francis. « L'irréalité de la mort », *In* : LENOIR, Frédéric et DE TONNAC, Jean-Philippe, (Eds.). *La Mort et l'Immortalité. Encyclopédie des Savoirs et des Croyances*, p. 1435-1436.

> impression musicale suivant qu'on les fasse entendre dans un ordre de succession ou dans l'ordre inverse.
>
> L'ordre de succession est donc essentiel. »

Nous n'évoquerons guère le problème de l'unité diachronique de la conscience dans le corps de notre thèse, mais nous en dirons quelques mots en annexes.

## 5. Fondements de l'identité numérique : (1) identité qualitative, ressemblance, connexité ; (2) continuité (qualitative et spatiale) et (3) identité numérique des constituants.

L'identité numérique est l'identité logique, c'est-à-dire l'identité qu'entretient une chose avec elle-même. Cette identité est nécessaire. Une chose ne peut pas cesser d'être identique à elle-même.

Des objets tels que les boules de billard, les chaises, les pierres, etc. peuvent rester pendant de très longues périodes de temps sans changer. Mais l'identité numérique n'est pas l'identité qualitative. On peut même douter que cette dernière en soit une condition nécessaire. Chez les êtres vivants, le changement, quoique le plus souvent progressif, est incessant. Pourtant, il semble intuitif de penser que tout au long de leur existence, ils forment une seule et même entité. Chez eux donc, il semble que la préservation de l'identité numérique tient à une ressemblance forte de proche en proche, à eux-mêmes, au cours du temps, et non à une identité qualitative stricte. C'est grâce à cette ressemblance de proche en proche que nous parvenons à correctement reconnaître un ami perdu de vue depuis un an comme celui qu'on a connu il y a un an.

Cette notion est proche de celle de connexité, développée par Parfit[40]. C'est-à-dire qu'il y a une certaine relation adéquate (que Parfit appelle « relation R »), le plus souvent une relation causale, qui relie chaque étape de la vie d'un individu à l'étape précédente et à l'étape suivante. Le maintien, au cours du temps, de cette relation, peut fonder l'identité numérique d'un individu au cours du temps.

Notons que l'on peut considérer l'identité qualitative stricte comme un cas particulier de relation R. C'est une relation R monotone.

Dans certains cas, l'élucidation de l'identité numérique passe par la connaissance de l'histoire et n'a rien à voir, de près ou de loin avec l'identité qualitative. Dans le cas d'une entité matérielle, c'est même, sans doute, son histoire qui constitue le critère le plus pur de son identité numérique. Par exemple, supposons que nous ayons une boule blanche dans notre main droite et une autre boule blanche indiscernable dans notre main gauche. Mélangeons ces deux boules dans un sac. Lorsque nous les reprenons, nous ne savons pas laquelle est laquelle, et nous ne pouvons pas le savoir. Mais si nous avions pu suivre leur histoire à chacune, voir quelles ont été leur trajectoire en passant de nos mains au sac, en se mélangeant dans le sac, puis en les ressortant du sac, nous aurions pu, très précisément, dire laquelle est la boule que nous tenions dans notre main droite et laquelle est la boule que nous tenions dans notre main gauche. Il n'y aurait eu aucune ambiguïté.

L'aspect historique de l'identité numérique est souvent négligé, à tort. Il est très important. Si un enfant a patiemment empilé des cubes, que son père, en l'absence de son fils, démoli par mégarde l'empilement de ce dernier et le reconstitue à l'identique avant que

40 Cf. PARFIT, Derek, "Reasons and Persons"

son fils ne s'aperçoive du méfait, la construction aura beau être indiscernable de celle d'avant, ce ne sera plus celle faite par l'enfant, mais une réplique faite par son père.

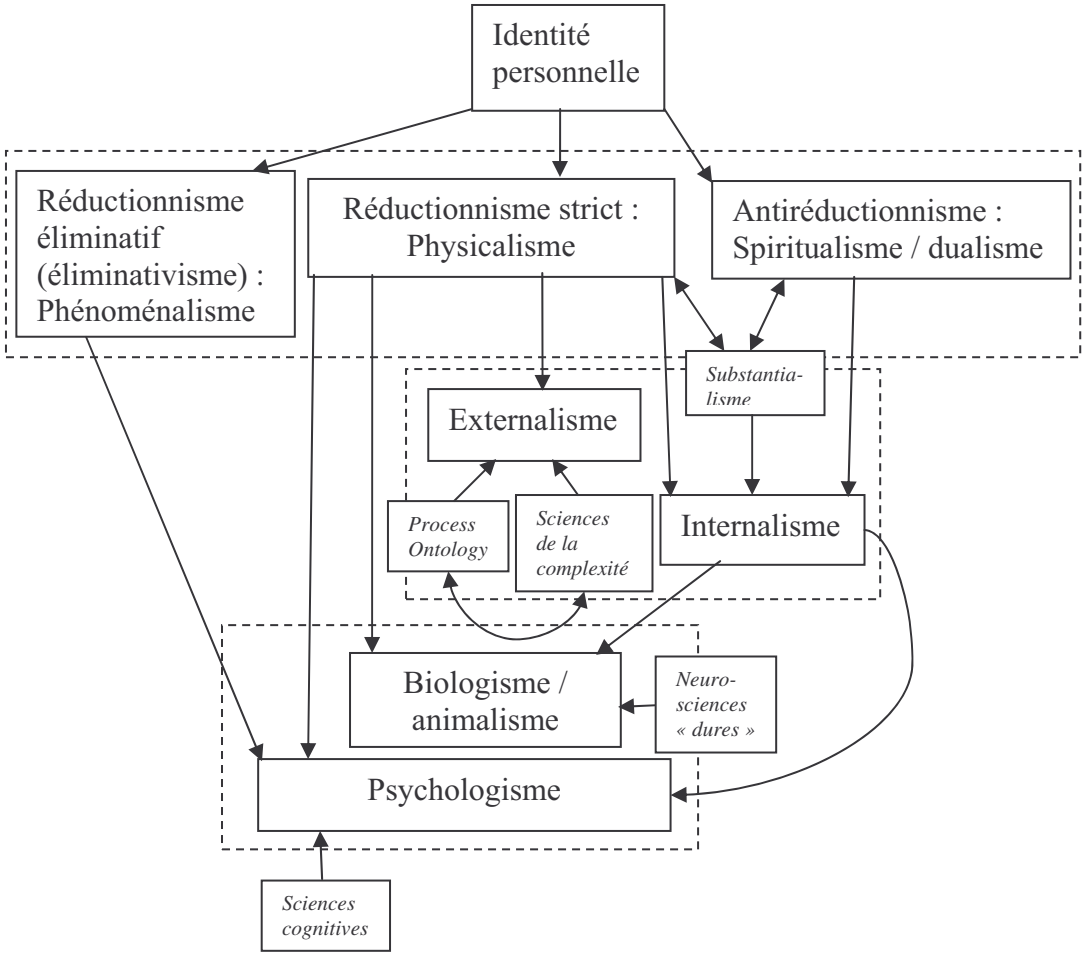
Dans certains cas, la considération de l'identité numérique des constituants d'une entité permettra de se prononcer sur l'identité numérique de l'entité. Par exemple, la cire d'une bougie obtenue en faisant fondre cette dernière lui est numériquement identique, pour la seule raison que les molécules de la cire fondue sont numériquement identiques aux molécules de la bougie avant sa fonte. Chaque molécule de la bougie a eu, en effet, une certaine trajectoire bien précise – laquelle pourrait en principe être reconstituée – qui fonde son identité numérique avec une molécule présente dans la cire fondue.

Le dernier problème que nous aurons à aborder concernant l'identité personnelle, consistera donc à définir quel type d'identité parmi toutes celles que nous avons cité, il convient de privilégier dans le cas des personnes.

### *B. Espace des solutions*

On peut, pour clarifier les choses, représenter l'espace des solutions aux problèmes soulevés par l'identité personnelle que les philosophes ont tenté d'apporter, en dessinant un petit schéma en arbre, qui permet de situer ces solutions les unes par rapport aux autres, en

matérialisant leur proximité conceptuelle et les liens catégoriels qu'elles entretiennent les unes avec les autres[41].




[41] **Nota bene :** les flèches matérialisent les rapports conceptuels entre les diverses théories et le sens en lequel s'effectue ce rapport ; les cadres pointillés regroupent les théories par grandes thématiques ; enfin, en petit et en italique, on trouvera quelques théories annexes, non directement liés à l'identité personnelle, mais dont on devine une certaine contribution indirecte, par le cadre qu'elles fournissent à la formulation de certaines théories sur l'identité personnelle.

## 1. Réductionnisme vs Antiréductionnisme

La première division conceptuelle oppose les théories antiréductionnistes de l'identité personnelle aux théories réductionnistes. Parmi ces dernières, l'éliminativisme est le plus radical.

Pour l'antiréductionniste, le concept d'identité personnelle se réfère à une réalité d'un certain type, le plus souvent de nature spirituelle. L'antiréductionniste prend en effet au sérieux la plupart des intuitions que nous avons justement qualifiées d'antiréductionnistes, concernant l'identité personnelle. L'antiréductionniste ne nie pas forcément une influence de la matière sur l'esprit, le premier pouvant conditionner l'expression du second. En revanche, il conteste que l'on puisse réduire le spirituel au matériel, le sujet unique, permanent et indivisible de l'aperception à la multiplicité changeante et corruptible de la matière. Le spiritualisme a longtemps dominé la pensée au sujet de l'identité personnelle, par force de tradition, et parce que la science n'en était qu'à ses balbutiements. De Platon à Descartes, et surtout à partir de ce dernier, c'est toute une tradition de pensée, qui se poursuit chez Kant, l'idéalisme allemand, puis les phénoménologues husserliens, avec parfois quelques variations. A notre époque, et malgré le renouveau de la pensée inspiré par les neurosciences, le spiritualisme influence encore peu ou prou les conceptualisations actuelles de la conscience. Ainsi, certains neuroscientifiques, dans l'esprit de La Mettrie, se sont donnés comme objectif de fonder un cartésianisme matérialiste, de trouver un soubassement neuronal au « théâtre cartésien »[42], ce lieu où prendrait place toutes nos représentations et images mentales, et qui expliquerait le sentiment d'unité de la conscience. On a donc cherché un lieu de convergence, dans le cerveau, de toutes les informations

[42] L'expression est de Daniel Dennett.

acheminées à la conscience. Selon une autre démarche, quelques rares philosophes et neuroscientifiques ont soutenus un dualisme explicite entre le cerveau et la conscience. C'est le cas, par exemple, de l'épistémologue Karl Popper, ou du prix Nobel de médecine Sir John C. Eccles[43].

A l'opposé de ces conceptions, une minorité de philosophes estime que le sentiment du moi est illusoire, qu'il n'y a pas, à proprement parler, d'identité personnelle. Actuellement, cette position est défendue par Derek Parfit[44]. On la trouve déjà formulée chez David Hume. Dans le *Traité de la Nature Humaine*[45], le philosophe écossais l'illustre par cette réflexion :

> « Pour ma part, quand je pénètre le plus intimement dans ce que j'appelle *moi*, je bute toujours sur une perception particulière ou sur une autre, de chaud ou de froid, de lumière ou d'ombre, d'amour ou de haine, de douleur ou de plaisir. Je ne peux jamais me saisir, *moi*, en aucun moment sans une perception et je ne peux rien observer que la perception. Quand mes perceptions sont écartées pour un temps, comme par un sommeil tranquille, aussi longtemps je n'ai plus conscience de *moi* et on peut dire vraiment que je n'existe pas. Si toutes mes perceptions étaient supprimées par la mort et que je ne puisse ni penser, ni sentir, ni voir, ni aimer, ni haïr après la dissolution de mon corps, je serais entièrement annihilé et je ne conçois pas ce qu'il faudrait de plus pour faire de moi un parfait néant. Si quelqu'un pense, après une réflexion sérieuse et impartiale, qu'il a, de *lui-même*, une connaissance différente, il me faut l'avouer, je ne peux raisonner plus longtemps avec lui. Tout ce que je peux lui accorder, c'est qu'il peut être dans le vrai aussi bien que moi et que nous différons essentiellement sur ce point. Peut-être peut-il percevoir quelque chose de simple et de continu qu'il appelle *lui* : et pourtant je suis sûr qu'il n'y a pas en moi de pareil principe. »

43 Cf. par ex.: POPPER, Karl Raimund, et ECCLES, John Carew, *The Self and its Brain.*

44 PARFIT, Derek. "Reasons and Persons".

45 HUME, David. *Traité de la Nature Humaine*, partie IV, section 6 : « De l'identité personnelle ».

Cette conception de l'être humain est cependant bien plus ancienne encore, puisqu'on la retrouve véhiculé par l'antique sagesse du Bouddhisme Theravada.

Selon cette conception éliminativiste, les cas jugés problématiques au regard de l'identité personnelle sont de faux problèmes, et nos intuitions antiréductionnistes peuvent être expliquées sans faire appel à un concept d'identité personnelle. Ce qu'il y a, tout simplement, c'est que nous sommes des individus relativement disjoints les uns des autres. Ainsi, nos champs d'expériences sont disjoints également : nous ne sommes conscients directement que de ce qui nous affecte, tandis que nous ne pouvons que deviner ou inférer ce qui affecte autrui. Cette clôture phénoménologique nous donne l'illusion d'être un être à part des autres, ontologiquement distinct, unique et indivisible. Comme, de plus, nous avons le souvenir vivace, en première personne, de notre passé, mais que nous ne pouvons pas nous remémorer, en première personne, les souvenirs qu'à autrui de son propre passé, nous imaginons un lien fort – ontologique – qui relie les différents moments de notre vie. Nous supposons que quelque chose d'identique à lui-même au cours du temps traverse les années et nous fait être le même, réellement, que l'enfant que nous pensons avoir été. Mais tout cela n'est qu'illusion également.

Enfin, selon une troisième conception de l'identité personnelle, il y a un véritable problème qui se pose et qu'il importe de résoudre. Mais pour le résoudre, il ne suffit pas de postuler une mystérieuse âme et d'affirmer que cette âme va résoudre tous nos problèmes, même si on ne sait pas comment… C'est là une hypothèse entièrement *ad hoc*. Ce qu'il faut faire, c'est rechercher, dans ce qui est donné à notre observation, quelque chose d'invariant, d'indivisible et d'unique qui fonde en raison notre identité personnelle. En postuler par

avance l'existence, comme dans le cas de l'âme, sans se donner les moyens d'une vérification (l'âme étant inobservable), c'est choisir la solution de facilité.

Deux stratégies principales vont être déployées dans cette quête d'un fondement objectif, empirique et rationnel de l'identité personnelle. Une stratégie va consister à explorer notre conscience de nous-même, nos souvenirs, notre personnalité, dans l'espoir d'en isoler l'essence de la personne. Une autre cherchera dans la matière même du corps (et notamment du cerveau), à fonder l'identité personnelle, en prenant modèle sur ce qui fonde l'identité numérique des objets en général (minéraux, roches, végétaux, animaux, artefacts, etc.).

## 2. Biologisme vs Psychologisme

De prime abord, lorsque l'on a laissé tombé peu ou prou l'idée de fonder l'identité personnelle sur une substance spirituelle, il semble tentant de conserver néanmoins à la vie mentale une certaine prééminence. Même si on admet, avec les neurosciences, une incarnation corporelle de l'esprit, on peut vouloir penser que c'est dans la mesure où nous conservons de mêmes pensées, que l'on observe une pérennité de nos souvenirs, une continuation de nos traits de personnalité, que nous sommes la même personne, pour soi comme pour les autres. Ce n'est pas parce que l'on conserve un même corps. Locke nous explique :

> « Pourrions-nous supposer deux conscience distinctes et incommunicables faisant agir le même corps, l'une de jour et l'autre de nuit, et en sens inverse la même conscience qui ferait agir par intervalle des corps distincts ? Je me demande si, dans le premier cas, *celui qui travaille de jour* et *celui qui travaille de nuit* ne seraient pas deux personnes

> aussi distinctes que *Socrate* et *Platon* ; et si dans le second cas, il n'y aurait pas une personne en deux corps différents, tout comme un homme reste le même dans des vêtements différents.
>
> Il ne sert absolument à rien de dire que dans les cas précédents les *consciences* sont les mêmes ou diffèrent en fonction de substances immatérielles identiques ou différentes, qui introduirait en même temps la conscience dans ces corps : que ce soit vrai ou faux, cela ne change rien, puisqu'il est évident que l'*identité personnelle* serait dans les deux cas déterminée par la conscience, qu'elle soit attachée à une substance immatérielle individuelle ou non. »[46]

Toutefois, très vite, des voix s'élevèrent à l'encontre de cette conception de l'identité personnelle. Des noms comme Leibniz, Thomas Reid et John Butler sont associés à cette critique des critères psychologiques de l'identité personnelle, ceux-ci étant jugés circulaires, non-transitifs, trop peu déterminants (comme nous le montrerons diverses expériences de pensée). Certains partisans des critères psychologiques tentèrent de répondre aux objections, et de peaufiner leur modèle. D'autres philosophes – assez minoritaires toutefois – eurent l'idée de s'intéresser au corps, ce grand oublié des philosophies idéalistes, pour y chercher un critère de l'identité personnelle. Chez la plupart des êtres, c'est en effet la matière, et non la fonction, qui assure, semble-t-il, le soubassement de leur identité. Une bougie que l'on fond n'a plus la même forme, la même consistance ni la même couleur. Pour autant, on s'accordera en général à considérer qu'il n'y a là qu'un changement de propriétés qui n'affecte en rien l'identité numérique de la bougie. La raison en est que l'on suppose l'existence d'un substrat matériel qui demeure identique tout au long du processus, quand bien même ses propriétés qualitatives se modifient. Ce sont les mêmes atomes qui constituent la bougie avant et après la fusion.

---

[46] LOCKE, John, *Essai sur l'entendement humain*, Livre II, chapitre XXVII : « Identité et différence », § 23, p 536-537.

Le recours aux critères somatiques dans la définition de l'identité personnelle consiste à comparer le cas de la personne à celui de la bougie dans notre exemple de tout à l'heure. C'est une conception compositionnelle de l'identité numérique. Un être demeure le même, si et seulement si ses constituants demeurent les mêmes. Une des premières formulations de cette approche peut être trouvé chez le philosophe grec Lucrèce, qui pourtant n'avait pas cherché à se pencher particulièrement sur le problème de l'identité personnelle.

Lucrèce écrivait :

> « Et quand bien même le temps, après notre mort, rassemblerait toute notre matière et la réorganiserait dans son ordre actuel en nous donnant une seconde fois la lumière de la vie, là encore il n'y aurait rien qui nous pût toucher, du moment que rupture se serait faite dans la chaîne de notre mémoire. Que nous importe aujourd'hui ce que nous fûmes autrefois ? que nous importe ce que le temps fera de notre substance ? En effet, tournons nos regards vers l'immensité du temps écoulé, songeons à la variété infinie des mouvements de la matière : nous concevrons aisément que nos éléments de formation actuelle se sont trouvés plus d'une fois déjà rangés dans le même ordre ; mais notre mémoire est incapable de ressaisir ces existences détruites, car dans l'intervalle la vie a été interrompue et tous les mouvements de la matière se sont égarés sans cohésion bien loin de nos sens. »[47]

Dans ce passage, Lucrèce expose ses conceptions matérialistes épicuriennes. On le sait, l'épicurisme, sous la plume de son fondateur, Epicure, professait que : « la mort n'est rien pour nous car lorsque nous sommes vivants, elle n'est pas là, et que lorsqu'elle vient, nous ne sommes plus ». Cette célèbre citation résonne encore chez beaucoup de nos contemporains qui font remarquer qu'il n'y a pas lieu de craindre la mort, parce qu'aucune

---

47 LUCRECE, *De la Nature des Choses*, Livre 3ème.

souffrance ne saurait atteindre une personne qui n'existe plus. La souffrance présuppose l'existence. La mort nous privant de cette dernière, nous délivre à jamais de la première.

Lucrèce, toutefois, envisage la possibilité d'une improbable résurrection qu'il explique comme suit. Supposons que les éléments qui nous constituent soient rassemblés à nouveau un jour, et dans le même ordre – il n'y aurait alors rien qui nous distinguerait de ce que nous étions auparavant – il semble que nous existerions à nouveau, de même que si l'on démonte un appareil et qu'on le remonte tel qu'avant, avec les mêmes pièces, on serait tenté de dire qu'on lui a redonné l'existence.

Si notre existence consistait dans le seul fait d'avoir certains éléments matériels dans un certain ordre, l'argument de Lucrèce serait valable. La biologie nous enseigne toutefois que les éléments matériels qui constituent notre corps sont dans un perpétuel recyclage. Si nous mangeons un animal, peu à peu, les éléments qui le constituaient seront intégrés à notre organisme. Inversement, lors des processus d'élimination : urine, fèces, sueur, métabolisme, etc. ce sont des éléments qui nous constituaient qui retournent à la nature.

L'hypothèse de Lucrèce peut cependant être maintenue, avec une petite modification mineure : ce qui est requit, dans la persistance de notre identité, ce n'est pas l'identité numérique stricte des éléments qui nous constituent, mais la continuité du flux d'assimilation – élimination de ces éléments.

S'identifier à son propre corps ne va pas forcément de soi. Pourtant n'est-ce pas d'abord grâce à leur corps, avant toutes choses, que nous reconnaissons nos semblables ? L'homme, avant d'être un être pensant, est d'abord – et essentiellement – un animal, selon

les partisans d'une théorie encore peu connue, l'*animalisme*, et cette essence englobe son essence de personne ou d'individu. D'autres, sans considérer que notre appartenance au règne animal est une condition nécessaire à la préservation de notre identité (nous sommes des animaux, mais nous aurions pu être des robots ou des anges), considèreront que nous sommes avant tout un corps.

Après Darwin, et son *Origine des Espèces*[48], que nous soyons des animaux ne pose plus guère de problème au grand public cultivé. Depuis Changeux et son *Homme Neuronal*[49], l'incarnation de l'esprit s'insinue progressivement dans les consciences. Toutefois, il reste une grande tentation cartésienne à faire de la pensée, de la conscience de soi, etc. (jugés comme ce qui fait notre spécificité d'homme), le siège de notre identité.

Le critère corporel, qui stipule ceci même que nous sommes notre corps, et que nous sommes la même personne, tout au long de notre carrière, parce que nous avons le même corps, n'inspire guère les philosophes contemporains, qui pour leur majorité adhèrent à un critère psychologique. Citons quand même, comme partisans de l'approche somatique : Wiggins[50], Thomson[51], Ayer[52], Inwagen[53], Olson, et, en France, Ferret[54] et Pouivet.

---

48 DARWIN, Charles. *L'Origine des Espèces.*

49 CHANGEUX, Jean-Pierre. *L'Homme Neuronal.*

50 WIGGINS, David. *Sameness and Substance.*

51 THOMSON, Judith Jarvis. "People and Their Bodies". *In:* DANCY, Jonathan (Ed.). *Reading Parfit.*

52 AYERS, Michael J., *Locke.*

53 VAN INWAGEN, Peter, "The Possibility of Resurrection".

54 FERRET, Stéphane, Le Philosophe et son Scalpel.

Le même corps, qu'est-ce à dire ? Notre corps, évidemment, change sans cesse, insensiblement. Le critère ne doit pas suggérer une identité qualitative stricte d'une phase à l'autre de notre développement, mais seulement une ressemblance de proche en proche, une continuité dans le processus de la vie qui nous anime et auquel participe, l'espace de quelque temps, un flux ininterrompu de molécules, d'énergie, que notre organisme assimile et rejette, pour maintenir un équilibre fragile qui assure son unité dynamique et sa persistance à travers le temps.

### 3. Internalisme vs externalisme

L'individu ne survivrait pas en isolation totale. L'environnement est essentiel au maintien de son équilibre. Il y trouve la nourriture dont il a besoin, les congénères avec lesquels il interagit, et qu'il reconnaît donc, d'abord par leur corps. Parmi eux, se trouve des compétiteurs dans la lutte pour la vie, mais aussi des partenaires pour l'accouplement et la reproduction, qui permet la poursuite de cette continuité du flux de la vie, par-delà la mort individuelle. L'environnement, qui a été encore plus négligé que le corps par les philosophes, a donc peut-être son importance dans la constitution de notre identité personnelle ou individuelle.

Les relations que nous établissons avec notre environnement font peut-être partie, de manière constitutive, de nous-même, et les critères somatiques ou psychologiques doivent peut-être être élargis à la prise en compte de cet environnement, dont on sait à quel point il est nécessaire à notre bien-être et peut influencer directement nos processus cognitifs, affectifs, comportementaux, etc.

Selon une hypothèse, connue sous le nom d'externalisme actif, d'externalisme véhiculaire ou encore d'hypothèse de l'esprit étendu, nous ne pensons pas seulement avec notre cerveau, mais avec tout notre corps et avec notre environnement. Cette thèse se marie bien avec une autre, défendue entre autres par Joël de Rosnay[55] et les membres du collectif *Principia Cybernetica*[56] (mais dont les prémisses remontent au moins à Teilhard de Chardin), suivant laquelle la Terre, avec tous ces hommes interagissant de manière riche, rapide et constante, surtout depuis l'émergence de l'Internet, forme une sorte de « pluricellulaire de l'esprit », un cerveau planétaire doué de conscience et dont nous sommes les neurones. La place que nous jouons dans ce vaste réseau nous déterminerait dans notre identité personnelle, à la manière dont chaque neurone de notre cerveau peut être dit déterminé dans son identité numérique, par la place qu'il occupe dans le réseau de neurones, quand bien même, qualitativement, rien ne le différencie de n'importe quel autre neurone.

---

55 DE ROSNAY, Joël, *Le Macroscope*.

56 HEYLIGHEN, Francis, JOSLYN, Cliff et TURCHIN, Valentin, *Principia Cybernetica Web* [en ligne] : < http://pespmc1.vub.ac.be/ >

# Partie 2 : Discussion et critique

« “[…]

— Je crois que c’est cela qui vous exaspère, non ? dit Helen. Que la plupart des gens s’obstinent à croire qu’il y a bel et bien un fantôme dans la machine, malgré tous les savants et les philosophes qui leur répètent qu’il n’y en a pas.

— Ce n’est pas exactement de l’exaspération, réplique Ralph.

— Mais si. On dirait que vous êtes résolu à éradiquer la foi de l’humanité tout entière. Comme un inquisiteur s’efforçant d’éradiquer l’hérésie.

— Je crois simplement que nous ne devrions pas confondre une vérité qui nous arrangerait avec ce qui est réellement la vérité.

— Mais vous admettez que nous avons des pensées intimes, secrètes, connues de nous seuls.

— Certes.

— Vous admettez que ce que je ressens en ce moment, en me prélassant dans l’eau chaude sous le ciel étoilé, n’est pas tout à fait identique à ce que vous ressentez ?

— Je vois où vous voulez en venir, réplique Ralph. Vous êtes en train de me dire : c’est quelque chose d’être vous, c’est autre chose d’être moi, il y a une appréhension du monde qui est unique pour chacun de nous, qui ne peut être décrite objectivement ni expliquée en termes purement physiques. Pourquoi donc ne pas la qualifier d’être immatériel ou d’âme ?

— Oui, c’est à peu près ça.

— Et moi, ce que j’affirme, c’est qu’il s’agit quand même d’une machine. Une machine virtuelle dans une machine biologique.

— Tout est donc une machine ?

— Tout ce qui traite de l’information, oui.

— Je trouve que c’est une idée effroyable.”

Il hausse les épaules et sourit. “Vous êtes une machine qui a été culturellement programmée pour ne pas reconnaître qu’elle est une machine.” »

(David Lodge, *Pensées secrètes*)

# I. Introduction

L'identité numérique, qu'elle concerne les personnes ou d'autres entités, est un concept problématique, déjà difficile à définir, puisque c'est sur lui que repose toutes les définitions[57]. Il existe un problème – voire des problèmes – de l'identité tout court[58], avant d'y avoir un problème – voire des problèmes – de l'identité personnelle plus spécifiquement. Les difficultés soulevées par le concept d'identité remontent au moins à Parménide et Héraclite. Comment concilier nos intuitions *a priori* sur les concepts d'être et de non-être, avec la nature changeante des phénomènes que l'on constate *a posteriori* ? Indépendamment de toute référence au concept de personne, on peut illustrer les problèmes relatifs à l'identité, en prenant l'exemple du bateau de Thésée,

> « bateau dont les matériaux sont progressivement remplacés au fur et à mesure des traversées de Thésée entre le Pirée et Délos et dont les sophistes d'Athènes se demandaient s'il s'agissait toujours du même

[57] Définir M comme étant *p*, cela revient à dire que le mot M et la proposition *p* ont le même sens, ou tout au moins renvoient au même référent, sont numériquement identiques. Par exemple : Définir – *a priori* – le Soleil comme l'astre du jour, revient à poser l'identité numérique : « Soleil » = « astre du jour ». Définir – *a posteriori* – l'eau comme étant la molécule de formule $H_2O$, revient à poser l'identité numérique : « eau » = « $H_2O$ ».

[58] L'usage des expériences de pensée abonde dans la littérature consacrée au problème de l'identité personnelle, ce qui est parfois critiqué (cf. partie 3 : « Identité personnelle et expériences de pensée »). Toutefois, puisqu'il existe un problème – voire des problèmes – de l'identité tout court, avant d'y avoir un problème – voire des problèmes – de l'identité personnelle, cela constitue un argument en faveur du recours aux cas fictifs dans l'élucidation de l'identité personnelle, lorsque l'on s'y intéresse à des aspects qui sont déjà problématiques au sujet de l'identité plus court – ce qui est généralement le cas, les problèmes sont juste plus saillants en ce qui concerne l'identité personnelle – car en effet, même si le concept de personne n'était plus applicable dans certains scénarios imaginaires très éloignés des cas actuels, le concept d'identité s'y appliquerait malgré tout.

> bateau. […] A partir du moment où le processus est progressif, il est facile de se persuader que le bateau entièrement recomposé et le bateau primitif sont le même bateau. Cependant, si les marins d'Athènes plutôt que de remplacer progressivement les matériaux du bateau de Thésée avaient construit, parallèlement au bateau de Thésée, un second bateau, personne ne pourrait soutenir que ce bateau est le bateau de Thésée en dépit du fait que l'on peut supposer que les planches de bois qui sont introduites dans le bateau de Thésée dans la première hypothèse sont numériquement les mêmes planches que celles qui composent ce second bateau dans cette seconde hypothèse. En ce cas, ce second bateau dans cette seconde hypothèse et le bateau de Thésée perpétuellement réparé dans la première hypothèse seraient un seul et même bateau puisqu'il s'agirait *ex hypothesi* de la même sorte de chose, de la même forme composée rigoureusement des mêmes matériaux. Comment dès lors pourrait-il s'agir de deux choses différentes ? Mais puisque personne ne pourrait croire que le bateau de la seconde hypothèse est identique au bateau originel de Thésée, comment admettre que le bateau complètement réparé de la première hypothèse puisse être identique à celui-ci ? »[59]

Toutefois, c'est avec l'identité personnelle que ces difficultés se font plus saillantes. On peut toujours admettre que les objets ordinaires sont illusoires, que l'on a juste des flux de sensations dont la cohérence nous fait croire abusivement en l'existence de choses persistantes. Ainsi, le problème évoqué par le bateau de Thésée peut toujours être jugé non réel. Le bateau de Thésée n'existe pas vraiment, ce n'est qu'un nom que l'on donne à un certain nombre de sensations coordonnées dans notre esprit. Donc cela n'a pas de sens de se demander lequel des deux bateaux est le vrai bateau de Thésée. Mais cela est beaucoup plus difficile à admettre concernant l'esprit justement, surtout lorsqu'il s'agit du nôtre. Il y a une résistance intuitive qui semble insurmontable, à le faire. En partie, sans doute, parce que c'est dans notre esprit que s'enracinent précisément nos intuitions sur l'identité numérique en général, parce que nier l'identité de notre esprit reviendrait à nier le concept

---

[59] FERRET, Stéphane, *L'identité*, p. 25-26.

d'identité numérique lui-même, sur lequel se fonde toute notre compréhension du monde, et aussi, parce que le concept même de changement ne prend sens que par rapport à l'identité. Une chose ne change que si elle reste essentiellement elle-même, sinon il s'agit de disparition pure et simple *in nihilo* et d'apparition pure et simple *ex nihilo*. Mais même les concepts de disparition et d'apparition présupposent malgré tout un référentiel ultime, l'univers, par rapport auquel ils prennent sens et au regard duquel ils constituent des changements (une montagne disparaît, c'est un changement de l'univers, un océan se forme, c'est un autre changement de l'univers, etc.). Lorsque quelque chose disparaît, il laisse quelque chose d'autre à sa place (si on enlève ce magasin, on voit la colline qui est derrière, si on enlève ce papier peint, on voit le mur qui est derrière, etc.), qui comble son absence. Lorsque quelque chose apparaît, il prend la place de quelque chose d'autre (si on met un magasin ici, on ne voit plus cette colline, si on met ce nouveau papier peint, on cache l'ancien, etc.). Toute chose n'apparaît ou ne disparaît que sur fond d'autres choses qui disparaissent ou apparaissent, respectivement. Si, donc, ce que nous prenons pour des objets ne sont que des flux de sensations qui surgissent et s'évanouissent avec une rapidité inconcevable, il semble nécessaire que quelque chose persiste, l'esprit qui en est le témoin. [60]

Si l'identité personnelle est un leurre, alors il faut être prêt à croire sincèrement que notre esprit n'existe qu'en cet instant précis, qu'il n'y a pas plus de ciment ontologique – mais seulement un ciment existentiel – entre deux instants de notre vie qu'entre un instant de notre vie et n'importe quel instant de la vie de n'importe qui d'autre. Il n'y aurait pas de

[60] Stéphane Ferret considère que les questions sur l'identité portent sur les choses, et pas seulement sur les concepts, lorsque les choses en question sont dotées d'un principe interne d'activité (ce qui est vrai des personnes et des êtres vivants, mais faux des pierres, fleuves ou des artefacts comme le bateau de Thésée). Cela semble douteux, dans la mesure où un simple thermostat possède un principe interne d'activité, alors que pourtant, les questions relatives à son identité semblent se rapprocher davantage du cas du bateau de Thésée que du cas des personnes.

sens à se demander, en voyant quelqu'un, s'il est la même personne ou non qu'une autre que nous avons connu autrefois, quant à la mort, elle ne serait pas différente pour nous, au fond, que l'égrenage incessant de chaque instant de notre vie, puisque c'est au passage de cet instant à l'instant suivant que nous mourrons. L'instant qui suit ou qui précède ne nous concerne en rien particulièrement – ou disons, ni plus ni moins que la vie de n'importe qui d'autre – et c'est maintenant que nous devons être heureux, pas après – car après, ce sera quelqu'un d'autre qui sera heureux à notre place – ce qui implique de se faire un maximum de plaisir immédiatement, sans égard pour les conséquences funestes que cela peut avoir à l'avenir. Si l'on tient à se soucier de son avenir et à se sentir responsable de son passé, alors l'attitude logique, en absence d'identité personnelle, consiste à se sentir responsable et à se soucier de tout un chacun – même le parfait inconnu qui vit à l'autre bout du monde – dans exactement la même mesure, ce qui est à l'opposé de l'attitude précédente mais paraît tout aussi délirant, quoique moins choquant.

Que l'on se sente contraint de croire en l'identité de la personne à travers le temps[61] est une chose. Une autre en est que de justifier cette croyance. Il faudrait parvenir, semble-t-il, à mettre en évidence un aspect de nous-même que nous serions le seul à posséder, et que nous posséderions toute notre vie durant. Un aspect si possible essentiel au regard de

[61] Au moins de notre personne, car les autres, dans une perspective solipsiste, peuvent encore, à la limite, être considérés comme de simples objets au même titre que les tables et les chaises.

ce qui constitue ou paraît constituer ce que nous sommes ici maintenant[62]. Au minimum, il nous faudrait établir une relation suffisamment forte, et exclusive, entre tous les instants de notre vie.

La plupart au moins des critères généralement avancés pour rendre compte de l'identité personnelle sont aporétiques. C'est-à-dire que, bien qu'ils soient efficaces dans un grand nombre de cas, ils finissent toujours par rencontrer un obstacle conceptuel, par montrer des faiblesses plus ou moins gênantes qui culminent en paradoxes qui sont de véritables impasses.

Face à ce constat, il est intéressant de remarquer qu'antiréductionnistes et éliminativistes, que tout semblerait opposer, se rejoignent pourtant sur un point essentiel : le problème de l'identité personnelle est insoluble. Seule diffère la raison de cette insolubilité. Pour les antiréductionnistes, le problème est insoluble car il est inaccessible à notre entendement. Pour les éliminativistes, il est insoluble car en quelque sorte… il n'existe pas ! Peut-on vraiment dire qu'il n'y a pas de problème de l'identité personnelle ? Nous ne prétendons pas, dans ce chapitre introductif, répondre définitivement à la question, mais nous essayerons de sensibiliser le lecteur au type de problèmes qui se posent à nous lorsque l'on examine les critères de l'identité personnelle, à travers l'exemple d'une

---

[62] Si nous naissions tous avec un code barre unique derrière la nuque, nous aurions trouvé un critère épistémologiquement satisfaisant de l'identité personnelle, car nous n'aurions qu'à regarder ce code barre pour déterminer avec précision qui nous sommes. Néanmoins, d'un point de vue ontologique, il semble difficile à imaginer qu'un simple code barre puisse constituer l'essentiel de ce que nous sommes. On verrait plutôt des choses telles que la conscience de soi, les souvenirs, le corps, etc. Ce code barre ne serait au mieux qu'un indice de reconnaissance, un symptôme de l'identité – fut-il fiable à 100% – mais non son fondement. Sinon, rajouter une barre à ce code barre équivaudrait à modifier l'identité d'un individu : faire disparaître une personne, en faire apparaître une autre qui lui ressemble de près.

expérience de pensée fameuse : celle de la division du cerveau. C'est une des meilleures qui soit, à plus d'un titre :

- Tout d'abord, elle se fonde sur des cas cliniques réels : pour soigner certaines formes sévères d'épilepsie, on eut ainsi recours à un sectionnement chirurgical du corps calleux qui relie les deux hémisphères cérébraux (commissurotomie).

- Ensuite, elle malmène sévèrement nos intuitions concernant l'unicité et l'indivisibilité de la conscience. Or, ces intuitions antiréductionnistes sont sans conteste celles qui sont les plus prégnantes, celles auxquelles on a le plus de mal à renoncer. Arriverait-on, en effet, à s'imaginer conscient simultanément de deux sensations tout en n'étant pas conscient d'en être simultanément conscient ?

- Enfin, la division du cerveau est l'une de ces expériences de pensée sur laquelle butent plusieurs conceptions de l'identité personnelle : critères somatiques, critères psychologiques, spiritualisme, aucun de ces modèles n'en sort indemne. Aucun ne parvient simplement à en rendre compte.

Pour Derek Parfit[63], la notion d'identité personnelle est une notion bien pratique dans la vie de tous les jours mais qui ne correspond à aucune réalité ontologique. En conséquence de quoi, la croyance selon laquelle « tous les cas impliquant la notion d'identité personnelle doivent avoir une réponse » est fausse. De toute façon, à supposer

63 Cf. PARFIT, Derek. "Personal Identity".

qu'elle renvoie à une réalité ontologique, celle-ci serait sans intérêt pour nous. Répondre à des questions importantes pour nous, telles les questions concernant la mémoire, la responsabilité morale, et même la survie n'exigerait pas l'introduction de considérations quant à l'identité personnelle. Ce qui est important pour nous dans la question de la survie, selon Derek Parfit, ce sont les relations de « continuité psychologique », de « connexité psychologique » et la conservation de la mémoire. Toutes ces relations étant des questions de degré, contrairement à la relation d'identité numérique qui fonctionne en mode « tout ou rien ». Cela implique que la question de la survie est une question de degré elle aussi, qu'on peut donc survivre un peu, beaucoup, pas du tout, à tel ou tel événement, ce qui n'est pas une notion très claire et est un peu contre-intuitif[64].

De plus, ces relations, contrairement à la relation d'identité, ne sont pas uniquement bijectives mais peuvent être également surjectives (cas de fissions) ou injectives (cas de fusions). A propos de cas de fissions, Derek Parfit, continuant dans sa logique, va même aller jusqu'à affirmer textuellement ce qui apparaît comme un paradoxe logique flagrant : « Nous pouvons suggérer que je survis en tant que deux personnes différentes sans que cela implique que je suis ces personnes. »

Comment imaginer que je puisse survivre en tant que deux individus différents ? Comment peut-on dire que je survis si je ne suis aucun des individus résultant de la

---

64 Peut-être faut-il comprendre ces degrés en terme d'attention. L'attention est un processus qui comporte assurément des degrés. Au niveau du focus attentionnel, elle est à son paroxysme, puis elle décroit de part et d'autre jusqu'à l'inconscience pure et simple. Dans une perspective cette fois-ci diachronique, l'attention varie également. Lorsque l'on est en période de sommeil profond, elle est nulle car il n'existe, semble-t-il, aucune conscience. A l'état de veille non vigilante et durant le sommeil paradoxal, on a une attention très faible, très diffuse. A l'état de veille dite vigilante, l'attention est bien plus importante, et elle atteint des sommets lorsque l'on se concentre sur une activité complexe.

fission ? Mais dans le cours de son exposé il ne donne aucune réelle raison de reconsidérer ce paradoxe. Ses arguments et explications tournent continuellement autour du pot sans jamais pointer du doigt le nœud du problème.

Ce nœud du problème c'est que la conception de Derek Parfit n'est, semble-t-il, défendable qu'en niant l'essence de la subjectivité, en ne considérant cette dernière que sous l'angle de son *contenu* phénoménologique, sans égard pour ce que, faute de mieux, nous appellerions son *contenant*, et qui se caractérise notamment par le caractère *unifié* (que pourtant, il ne semble pas nier) qu'il semble imposer au contenu.

Concrètement, lorsque Derek Parfit décrit une expérience de pensée au cours de laquelle il sépare ou réunit à volonté ses hémisphères cérébraux afin, par exemple de résoudre un problème de mathématique de deux façons différentes en parallèle et qu'il écrit :

> « [17] Quand je déconnecte mes hémisphères, ma conscience se divise en deux courants. **Mais cette division n'est pas quelque chose que j'expérimente.** Chacun de mes deux courants de conscience semble avoir été franchement continu avec mon unique courant de conscience antérieur au moment de la division. Le seul changement dans chaque courant de conscience est la disparition de la moitié de mon champ visuel et la perte de sensation et de contrôle moteur dans la moitié de mon corps. » [65]

On est en droit de se poser les questions suivantes : si la division n'est pas quelque chose que j'expérimente (ce qu'il admet et est logique compte tenu du caractère unifié de la subjectivité), alors il ne semble pas y avoir d'autres possibilités que les suivantes :

---

65 Cf. PARFIT, Derek. "Personal Identity".

(1) J'expérimente de manière intégrée les informations des deux hémisphères, mais cela semble impossible à moins de supposer une sorte de télépathie constante entre hémisphères

(2) J'expérimente de manière intégrée les informations d'un seul hémisphère (mais lequel ?)

(3) Je n'expérimente plus rien. Si « Je » désigne un locuteur dont les hémisphères cérébraux sont reliés, alors je disparais nécessairement lorsque mon cerveau est divisé, et donc je n'expérimente rien. Mais cela semble être une manière un peu trop restrictive de me définir.

(4) J'expérimente chaque hémisphère l'un après l'autre, dans une sorte de fonctionnement en multitâche préemptif, mais alors je n'en ai pas conscience car, quand je passe d'un hémisphère à l'autre je n'emporte pas d'information, sinon on serait dans un cas très similaire à (1), et donc tout se passe subjectivement comme si je n'expérimentais qu'un seul hémisphère, ce qui est un cas très similaire à (2).

(5) J'expérimente chaque hémisphère en même temps dans une sorte de fonctionnement en multitâche coopératif, mais alors je n'en ai pas conscience puisque aucun hémisphère n'a accès aux informations situées dans l'autre, sinon on serait dans un cas très similaire à (1), et donc tout se passe subjectivement comme si je n'expérimentais qu'un seul hémisphère, ce qui est un cas très similaire à (2).

On le voit clairement, seules les solutions (1) et (2) sont viables, mais elles posent problème : la solution (1) impliquerait l'existence d'une télépathie constante entre hémisphères et la solution (2) entraînerait un indéterminisme subjectif.

Les problèmes logiques apparaissent d'une manière encore plus vive si on se pose, lorsque notre esprit n'est pas divisé, quelques questions sur notre état futur une fois divisé. Prenons un exemple. Supposons la situation suivante : on décide de séparer vos hémisphères cérébraux et on vous explique que l'on conditionnera votre hémisphère droit à aimer le chou et le gauche à le détester. Pour situer les choses, admettons que l'ensemble du processus prenne 24 heures. Quelle est la valeur de vérité de l'affirmation suivante : « Dans 24 heures, vous aimerez le chou. ». Examinons les possibilités, a priori :

- L'affirmation est vraie, pourtant votre conscience de l'hémisphère gauche ne l'aimera pas, or cette conscience fait partie de vous au même titre que l'autre.

- L'affirmation est fausse, pourtant votre conscience de l'hémisphère droit l'aimera, or cette conscience fait partie de vous au même titre que l'autre.

- L'affirmation est vraie et fausse à la fois, mais comment cela pourrait-il être possible ? On pourrait vouloir répondre que l'affirmation sera vraie du point de vue de la conscience de l'hémisphère droit et fausse du point de vue de la conscience de l'hémisphère gauche, mais une telle réponse ne répond pas à la question posée. En effet, ce ne sont

pas des consciences hémisphériques considérées séparément dont il est question mais de vous comme un tout. Or, vous ne pourrez pas avoir simultanément le sentiment d'aimer le chou et en même temps le sentiment de le détester. Il est vrai que certaines pathologies psychiatriques, comme la schizophrénie, sont caractérisées par la cohabitation de pensées contradictoires, mais si le fait de séparer vos hémisphères avait pour résultat de vous rendre schizophrène, alors nous serions tous schizophrènes du simple fait de l'existence des autres gens. En effet, la situation de séparation spatiale de chaque cerveau relativement à un autre cerveau n'est guère différente de la situation de chaque hémisphère d'un cerveau par rapport à l'autre hémisphère, quand les hémisphères ont été séparés.

- L'affirmation n'est ni vraie ni fausse, elle n'a pas de valeur de vérité. Or, puisqu'il semble évident qu'il y a bien quelque chose qui aimera le chou (la conscience de l'hémisphère droit) et quelque chose qui n'aimera pas le chou (la conscience de l'hémisphère gauche), cette situation n'est possible que si aucune de ces consciences n'est vous ou une partie de vous. Ce qui reviendrait à dire, semble-t-il, que vous n'existeriez tout bonnement plus.

Or, Derek Parfit refuse l'éventualité qu'une séparation des hémisphères doive être considérée comme un décès. Il le refuse pour de bonnes raisons :

« [9] Wiggins alors imagina sa propre opération. Mon cerveau est divisé, et chaque moitié est logée dans un nouveau corps. Les deux

personnes résultantes ont mon caractère et les mémoires apparentes de ma vie. Que m'arrive-t-il ? Il semble y avoir seulement trois possibilités : (1) Je ne survis pas ; (2) Je survis en tant qu'une de ces personnes ; (3) Je survis en tant que les deux personnes.

[10] Le trouble avec (1) est celui-ci. Nous sommes d'accord que je pourrais survivre si mon cerveau était transplanté avec succès. Et des gens ont dans les faits survécu avec la moitié de leur cerveau détruit. Il semble s'ensuivre que je pourrais survivre si la moitié de mon cerveau était transplantée avec succès et l'autre moitié était détruite. Mais si c'est ainsi, comment ne pourrais-je ne pas survivre si l'autre moitié était aussi transplantée avec succès ? Comment un double succès pourrait-il être un échec ?

[…]

Un homme qui regarderait l'opération de Wiggins comme une mort doit, je suggère, être ouvert à une de ces critiques. Il pourrait croire que sa relation à chacune des personnes résultantes échoue à contenir un certain élément qui est contenu dans la survie. Mais comment cela pourrait-il être vrai ? Nous sommes d'accord qu'il *survivrait* s'il se tenait dans cette exacte relation envers une seule des personnes résultantes. Aussi, cela ne peut être la nature de cette relation qui la fait échouer, dans le cas de Wiggins, à être une survie. Cela peut être seulement sa duplication.

[…]

[32] Pour ceux qui pensent que la question de l'identité est juste une question de décision personnelle, ce serait clairement absurde de regarder l'opération de Wiggins comme une mort. Ces gens devraient penser, "Nous pourrions avoir choisi de dire que je devrais être une des personnes résultantes. Si nous l'avions fait, je n'aurais pas dû la regarder comme une mort. Mais puisque nous avons choisi de dire que je ne suis aucune de ces personnes, alors tel est le cas."

C'est même dur à comprendre. »[66]

## Et pour de moins bonnes… :

« [30] Supposons qu'[un] homme [qui accepte que ce ne puisse pas être la nature de la relation qui la fait échouer, dans le cas de Wiggins, à

[66] Cf. PARFIT, Derek. "Personal Identity".

> être une survie, mais sa duplication], regarde encore la division comme une mort. Sa réaction semblerait maintenant follement inconsistante. Il serait comme un homme qui, alors qu'on lui parlerait d'un médicament qui puisse doubler ces années de vie, regarde la prise de ce médicament comme mortelle. La seule différence dans le cas de la division est que les années supplémentaires s'écoulent simultanément. C'est une différence intéressante. Mais cela ne signifie pas qu'il n'y a *aucune* année qui s'écoule. »[67]

La comparaison ne tient pas. Ce n'est pas la seule différence entre le cas de la division et le cas du doublement des années de vie. En effet, dans ce dernier cas, on peut supposer raisonnablement que lorsque l'homme en question commencera à vivre ses années supplémentaires, il ne perdra pas soudainement la mémoire de toute sa vie d'avant. Or, dans l'expérience de Wiggins, il n'y aura aucune information qui passera d'un individu résultant à l'autre, à moins que les individus en question se rencontrent et commencent à parler d'eux-mêmes, mais l'information échangée ne sera pas reçue par son destinataire, comme si elle correspondait à quelque chose vécu par lui. Du reste, une telle information, transmise uniquement par des mots, au mieux des films et des photos, sera très pauvre comparée à un authentique souvenir.

Une autre conceptualisation moins problématique de la théorie de Parfit est peut-être possible. Au lieu de nier l'identité personnelle, on peut aussi la démultiplier : chaque instant de notre vie verrait donc mourir un ancien « moi » et naître un nouveau « moi » numériquement distinct, et il n'y aurait aucun lien réel particulier entre ces différents « moi » qui se succèderaient les uns aux autres à chaque moment de notre existence. Cette vision des choses est qualifiée de *ponctualisme*, par Mark Brown[68]. De prime abord, cela

---

67 Cf. PARFIT, Derek. "Personal Identity".

68 Cf. BROWN, Mark. "Two Advantages of Punctualist Theories of Personal Identity".

semble profondément contre-intuitif. Il ne s'agit rien de moins que de nier qu'une quelconque existence de notre conscience soit possible plus d'un instant. Il s'agit d'affirmer que nous n'existons que de manière instantanée, maintenant, jamais avant et plus jamais après. L'impression qu'on aurait d'avoir existé avant serait due au souvenir qu'on aurait d'avoir existé avant, souvenir qui serait en fait trompeur, car ce passé aurait été vécu à chaque fois par un « moi » différent, une conscience différente. Il n'y aurait rien de réel qui unirait ces différents « moi » instantanés et qui permettrait de parler d'existence dans la durée.

Une telle conception semble assez proche de la notion bouddhiste d'*anatman* (ou *anatta*), non-soi, qui ne signifie pas à proprement parler que « je » n'existerais pas, mais que « je » n'existerais pas de façon indépendante et permanente. Il s'agirait d'une succession d'instants de conscience qui seraient produits par une cause dans des conditions données. Que ces phénomènes de conscience s'inscrivent dans une continuité, et nous croyons en la permanence d'un phénomène unique qui serait le « moi » à proprement parler.

Une certaine lecture de David Hume, notamment le passage, déjà cité[69], où il s'évertue à rechercher un « moi » permanent, qu'il dit ne pas trouver, à travers ses expériences d'introspection, suggère de semblables vues. Pour le philosophe écossais, ce qu'on appelle « Moi » ne serait qu'un terme qu'on attribue par convention à un paquet de perceptions.

Il est assez difficile de critiquer une telle vision des choses. Elle esquive toute aporie, il est vrai, mais elle présente le défaut d'un caractère profondément contre-intuitif. Bien sûr,

---

69 HUME, David. *Traité de la Nature Humaine*, partie IV, section 6 : « De l'identité personnelle ».

des choses peuvent être vraies et contre-intuitives mais en ce cas il est légitime de souhaiter qu'elles soient à tout le moins solidement argumentées… Or l'existence d'un « moi » un tant soit peu permanent est aussi sûre que celle d'un monde extérieur car à vrai dire, c'est un processus sensiblement semblable qui nous donne la conviction de l'existence d'un « moi » et d'un monde extérieur, à savoir la cohérence de nos perceptions à travers le temps, dont on prend conscience à travers notre mémoire.

Après ce petit chapitre introductif, nous sommes armés pour attaquer le gros du développement concernant le problème de l'identité personnelle. Le sujet est si vaste que nous nous sommes concentrés plus particulièrement sur le problème de l'identité personnelle à travers le temps[70]. Dans une première sous-partie, nous ferons dialoguer les arguments et contre-arguments des tenants de l'antiréductionnisme et des tenants du réductionnisme – notamment l'éliminativiste – en matière d'identité personnelle. Nos intuitions peuvent-elles trouver un appui rationnel ? En second lieu, nous évaluerons deux importants critères de l'identité personnelle : le critère psychologique et le critère somatique. Enfin, nous montrerons, dans une dernière sous-partie, la pertinence et l'intérêt d'une approche anti-individualiste (c'est-à-dire externaliste) de l'identité personnelle.

## II. Les intuitions antiréductionnistes sur l'identité personnelle

Notre étude débutera par une analyse critique des intuitions conduisant à considérer qu'il y a vraiment quelque chose de mystérieux – autrement dit qu'il y a un véritable

---

[70] Les personnes intéressées par le problème de l'identité personnelle à travers les mondes possibles pourront lire avec profit le petit livre de Ferret, Stéphane (1993), *Le philosophe et son Scalpel*, Ed. de Minuit, Paris, .notamment le chapitre 2 : « Identité personnelle et mondes possibles », p. 33-47.

problème à résoudre – au sujet de l'identité personnelle. Nous passerons donc au peigne fin de l'analyse toutes ces intuitions, en montrant tout d'abord sur quoi elles reposent, et ensuite ce qu'il convient d'en penser.

Avant de présenter ces intuitions, nous présenterons une expérience de pensée proposée par David Banach, fortement pourvoyeuse d'intuitions antiréductionnistes, et qui nous servira de base de discussion :

> « Imaginez que vous êtes enlevé par un quelconque tortionnaire tortueux et vous réveillez pour vous trouver incapable de voir le monde de façon normale, à travers vos yeux et vos sens corporels. A la place, votre seule perception est une scène visuelle composée de cinq personnes alignées contre un mur, chacune avec des fils et des électrodes variés menant à leur tête. Aucune de ces personnes ne vous semble le moindrement familière. Vous entendez une voix qui vous informe qu'une des personnes que vous voyez est **vous**, est le corps qui vous maintient en vie, et a le cerveau dans lequel vos pensées prennent place. (Evidemment, votre ravisseur a aussi effacé la mémoire de votre corps ou l'a changé puisque aucun des corps ne vous semble familier.) La voix vous informe qu'elle sera heureuse de vous fournir toute information objective que vous voudrez sur le monde. Elle vous permettra d'examiner les corps de plus près et d'observer des enregistrements de leur histoire, ou d'examiner la micro-structure de leur cerveau. Elle vous donnera toutes les informations que vous voudrez, excepté celles qui impliqueraient nécessairement une vue ou une perspective qui serait connectée de manière évidente à un de ces corps que vous voyez. Par conséquent, vous ne pouvez pas voir le monde à travers vos yeux ou sens corporels. La voix vous pose alors la question évidente : "Bien, qui pensez-vous que **vous** êtes ?" et je soutiens qu'il n'y a aucun moyen de dire lequel de ces corps vous êtes. Et ceci, pour les [raisons suivantes] :
>
> > (1) Il n'y a pas de connexion nécessaire entre la vision ou perspective que vous avez (et qui définit votre identité) et quelque propriété objective que ce soit. Par conséquent, vous pourriez noter des corrélations entre les événements dans le cerveau d'une des

personnes et vos pensées, mais il n'y a pas de raisons de supposer que cet événement dans le cerveau est vous, est ce qui a votre perspective. (Votre ravisseur peut avoir lié votre cerveau aux autres cerveaux si bien qu'ils sont tous corrélés.)

(2) Chaque caractéristique objective que vous avez peut être dupliquée (si nous supposons que notre ravisseur tortueux est assez intelligent). Ce sont des propriétés universelles. Chaque caractéristique que vous pourriez identifier comme montrant qui vous êtes pourrait être dupliquée dans tous les corps par votre ravisseur.

Bien sûr, il y a divers stratagèmes que l'on pourrait essayer pour déterminer qui on est :

**(1) Le stratagème de l'action :** Supposons que vous convainquiez votre ravisseur de vous permettre de regagner dans une certaine mesure un contrôle sur votre corps. Vous pourriez penser que vous pourrez déterminer qui vous êtes en observant qui bouge de la manière dont vous l'avez décidé. Disons que vous décidiez de bouger votre bras. Vous surveillez, avec anticipation, pour découvrir qui vous êtes, et, à votre surprise, aucune des personnes ne bouge leur bras, mais toutes bougent leurs jambes. Il n'y a pas de connexion nécessaire entre vos actes de volonté et leurs résultats objectivement observables. Et même s'il y en avait, votre ravisseur pourrait reproduire ces résultats objectifs dans les cinq corps. Il faut noter que la force de l'argument ne dépend pas de l'absence de connexion nécessaire entre les propriétés physiques et mentales […]. (Rappelez-vous que j'utilise « objectif » pour me rapporter à n'importe quel contenu de l'expérience, contenus privés y compris).

**(2) Le stratagème du caractère subjectif :** Imaginez que votre ravisseur vous autorise à pénétrer à l'intérieur du caractère subjectif de l'expérience associé avec chaque corps (par un procédé quelconque). Vous êtes au comble de la joie, pensant que vous allez sûrement être capable de dire qui vous êtes par comparaison du caractère subjectif de chaque corps avec ce que vous sentez. A votre stupéfaction, vous découvrez qu'aucun des corps n'a un caractère subjectif indiscernable du vôtre [En note : « On a seulement besoin

> d'imaginer que votre ravisseur dispose d'un moyen de vous cacher une certaine partie du caractère subjectif de votre expérience. »]. (Votre ravisseur a évidemment altéré la manière dont votre expérience subjective naît). Comme si ce n'était pas assez méchant, les personnes trois et quatre sont indiscernables. Le caractère subjectif de notre expérience implique aussi des propriétés qui sont seulement accidentellement connectées à notre perspective et qui sont universelles et peuvent être dupliquées. »

Pour faciliter notre discours, nous supposerons que vous êtes la personne A1. Vous contemplez, en face de vous, 5 personnes, que nous appellerons : A11, A12, A13, A14 et A15. Pour renforcer encore le pouvoir troublant de l'expérience de pensée, nous supposerons que toutes ces personnes sont indiscernables l'une de l'autre. Toutefois, semble-t-il, l'une de ces personnes est vous, A1, est le corps qui vous donne vie. Mais lequel ?

On pourrait se dire que, même si la réponse nous est inaccessible, elle serait en principe accessible pour quelqu'un qui aurait observé la manipulation de votre ravisseur. Quelqu'un, A1, a été kidnappé, puis on a déplacé son corps parmi d'autres corps indiscernables, et l'observation de cette mise en place permettrait de retrouver le corps original. Se serait donc bien une information objective qui déterminerait notre identité personnelle. On peut, cependant, compliquer un peu le scénario : Votre ravisseur pourrait disposer d'une technologie lui permettant de diviser les êtres humains comme des amibes, de manière à obtenir deux êtres humains parfaitement indiscernables de l'original jusque dans sa structure fine (donc ayant des souvenirs et une psychologie parfaitement indiscernables, du moins au début bien entendu).

A partir de A1, on aurait obtenu cinq personnes dont aucune ne possèderait un corps numériquement identique à celui de A1, mais toutes en auraient hérité d'une partie. La question à se poser, dans ce cas là, ne serait pas : Quel est le corps qui est celui de A1, mais, tout d'abord : A1 est-il encore en vie ? A-t-il survécu à ce processus ? Et si oui, alors seulement se demander : Où est-il ? Quel est le corps qui est son nouveau corps ? Ou, en se plaçant dans votre perspective : « Où suis-je ? Quel est le corps qui est mon nouveau corps ? ».

Si on se place dans le cadre de ce second scénario modifié, plusieurs solutions de prime abord rationnelles sont possibles :

- Tous les individus (A11, A12, A13, A14 et A15) sont A1, dans la mesure où ils sont indiscernables de A1, et engendré par lui par fission.

- Aucun des individus n'est A1. Ce processus de « division amibienne » a fait « perdre » A1 en cours de route, ou encore, le fait même de l'existence de plusieurs « candidats » tout aussi plausibles, pour être A1, fait qu'aucun de ces « candidats » n'est A1.

- Un et un seul des individus est A1.

La première solution est généralement écartée comme impossible par les anti-réductionnistes car elle impliquerait que A1 puisse se sentir exister à cinq endroits à la fois, en tant que cinq individus spatialement disjoints. Des réductionnistes éliminativistes comme Parfit n'y voient cependant aucun problème.

Réductionnistes éliminativistes et anti-réductionnistes s'accordent généralement pour rejeter la seconde solution, car on ne voit pas pourquoi l'existence de plusieurs « bons candidats » pour être A1 impliquerait qu'aucun ne soit A1, alors qu'il semble logique que si un seul bon candidat existait, il serait A1.[71] Certains partisans d'une forme de détermination externe de l'identité personnelle, tels David Wiggins et Roger Pouivet, trouvent néanmoins cette solution satisfaisante. Nous pensons également que, dans une certaine mesure, l'identité personnelle est déterminée par des facteurs externes, et pas uniquement par des facteurs internes, toutefois, nous ne trouvons pas cette solution satisfaisante, pour des raisons que nous expliciterons dans le chapitre 3 (« Des critères aporétiques ») de cette partie.

Il reste donc la troisième solution, qu'adoptent généralement les anti-réductionnistes – et rejettent quasi-unanimement les réductionnistes éliminativistes – et qui semble impliquer que, puisque les cinq individus résultants sont indiscernables de A1 et puisque seul un de ces individus est A1, alors l'identité personnelle est déterminé par autre chose que ce qui est publiquement observable et expérimentalement répliquable.

### *A. Intuition n°1 : Le fait manquant.*

*« Dans le cas où plusieurs individus futurs sont en continuité avec moi ou instancient les mêmes caractéristiques objectives que moi, l'impossibilité de dire lequel de ces individus je serai est dû à un manque d'information sur ce qui constitue mon identité personnelle ».*

---

[71] Nous discuterons de ce point plus en avant dans notre thèse, notamment dans le chapitre 3 de cette partie : « Des critères aporétiques » et dans la partie 3 : « Identité personnelle et expériences de pensée ».

Critique :

Cette impossibilité pourrait également être dû à un manque d'information au sujet d'un futur forcément incertain. Pour preuve, on peut montrer que le même genre de « problème » se poserait avec un scénario où il n'y aurait constamment qu'un seul individu.

Comparons les deux scénarios suivants :

1er scénario :

Imaginons que l'on dise à A1 qu'il va être dupliqué, simultanément, en 5 individus qui vont se retrouver à cinq points différents du globe : le Sahara, l'Arctique, l'Amazonie, le Tibet, l'Australie.

Si on demande à A1 à quel endroit il va se retrouver, il lui est impossible de répondre. C'est également une tâche impossible pour qui que ce soit d'autre. On pourrait penser que c'est parce que l'identité personnelle de A1 est déterminée par quelque chose, un certain fait, non objectif et non copiable, qui échappe à l'investigation empirique. Toutefois, examinons à présent le scénario n° 2.

2ème scénario :

Imaginons que l'on dise à A1 qu'un dé à 5 faces va être lancé et que, selon le résultat, il va être envoyé :

- ou bien au Sahara

- ou bien en Arctique

- ou bien en Amazonie

- ou bien au Tibet

- ou bien en Australie

Si on demande à A1 à quel endroit il va se retrouver, il lui est impossible de répondre. C'est également une tâche impossible pour qui que ce soit d'autre. Mais le fait qu'il soit impossible de répondre à la question n'implique pas qu'il nous manque des informations sur l'identité personnelle de A1. Ce qui nous manque, c'est une information sur un événement futur : le résultat du lancer de dé.

Par conséquent, il peut exister une indétermination quant à la destination de A1, dans le premier scénario, sans que cette indétermination soit liée à un manque de données sur l'identité personnelle de A1.

## *B. Intuition n°2 : Le « moi » fixe et permanent.*

*« Le "moi" est quelque chose de fixe une bonne fois pour toute (une essence) et de permanent au cours de la vie (le "moi" endure) »*

Critique :

Si le « moi » est une essence, c'est une essence indéfinissable car toute personne change énormément au cours de sa vie, et il paraît impossible d'isoler quelque chose de suffisamment stable qui lui serve d'assise. De plus, ce qui, en l'homme, semble le plus stable (ses gènes par exemple), peut être commun à plusieurs individus (dans le cas des jumeaux monozygotes, les gènes sont identiques) et/ou ne semblent pas avoir de lien avec nos intuitions au sujet du « moi » (les gènes, par exemple, ne sont pas conscients). Il semble donc, plutôt, que le « moi » n'est pas une essence et qu'il se redéfinit sans cesse au cours de la vie. Il se peut même qu'il disparaisse et soit remplacé par un autre « moi » (c'est possiblement le cas lors d'amnésies sévères).

Enfin, il semble que l'on puisse rendre compte d'une permanence du « moi » qui ne passe pas par la caractérisation d'une essence, mais par des critères exclusivement existentiels. Comme par exemple, la continuité dans l'évolution de certains traits physiques et psychologiques, la continuité dans notre localisation dans l'espace (nous ne disparaissons pas d'une place pour réapparaître peu après à une autre place, notre évolution dans l'espace se fait de manière continue, de proche en proche).

### *C. Intuition n° 3 : Le « moi » interne.*

*« L'identité du "moi" est entièrement déterminée de l'intérieur, par des facteurs et caractéristiques internes : mémoire, continuité psychologique, morphologique ou corporelle, « âme dans la tête », etc. »*

Critique :

On verra plus loin, avec l'étude de la thèse de l'externalisme actif, qu'il n'y a pas de barrière magique autour du cerveau qui délimiterait un intérieur de l'esprit, réellement mental, à l'intérieur du cerveau, face à un extérieur de l'esprit, seulement instrumentalement mental, dans le corps et l'environnement. Il n'y a donc pas de raison de supposer que l'identité du « moi » est entièrement déterminée de l'intérieur.

Il y a de bonnes raisons de penser que l'identité du « moi » est déterminée également par :

- les faits spatialement extérieurs au cerveau (i.e. les faits de l'environnement physique et social et les faits corporels) ;

- les faits passés voire à venir concernant : le cerveau, le corps, l'environnement physique et social.

Autrement dit, nous sommes, en partie, une certaine ligne dans l'espace-temps. Cette ligne contribue à définir notre identité.

Une telle stratégie permet de comprendre pourquoi deux individus indiscernables mais spatialement disjoints n'ont pas le même « moi ».

## *D. Intuition n°4 : J'aurais pu être n'importe qui d'autre.*

*« On peut s'imaginer être quelqu'un d'autre. On peut s'imaginer qu'au lieu d'être né là où on est né, et des parents que l'on a eu, d'avoir fait ce que l'on a fait et d'avoir eu ce que l'on a eu, que l'on soit né*

*ailleurs, à une autre époque, d'autres parents, d'avoir été différents, d'avoir fait des choses différentes, etc. Par exemple, nous aurions pu être un pharaon en Egypte il y a plusieurs millénaires, être Emmanuel Kant, la voisine d'à côté, Jacques Chirac, un extraterrestre, ou Mère Teresa.*

*Je suis donc quelque chose d'autre que tous les corps humains du monde, que tous les cerveaux, que toutes les personnalités et les souvenirs de qui que ce soit. Je n'expérimente jamais qu'un seul corps, qu'un seul cerveau, qu'une personnalité et que les souvenirs d'un seul individu, alors qu'il y en a des milliards d'autres que j'aurais pu expérimenter. »*

Critique :

Pourquoi est-ce bien moi qui suis moi ? Cette question, que se pose fréquemment les enfants, que se pose aussi Stéphane Ferret (même s'il se refuse à y répondre, arguant qu'on ne peut rien en dire), peut sembler vide de sens, aboutissant soit à une tautologie, soit à une contradiction. Car en effet, que je sois moi et pas un autre, n'est-ce pas une nécessité logique ? Et l'idée selon laquelle j'aurais pu être quelqu'un d'autre, une absurdité ? (car si j'avais été quelqu'un d'autre, je n'aurais pas été moi, mais le quelqu'un d'autre en question). Posé ainsi, le problème est effectivement un faux problème. La relation A=A n'apporte aucune information que nous ne connaissions déjà (c'est une proposition *a priori* mais elle n'est ni analytique ni synthétique). La relation A≠A est nécessairement fausse. On peut néanmoins donner sens au problème en ayant soin de distinguer l'aspect subjectif du concept de personne (*moi, je*), et son aspect objectif (*cette* personne). Le problème, correctement formulé, devient ainsi : Comment se fait-il que *je* sois *cette* personne ? Il y aurait donc deux sens en jeu : un sens subjectif et un sens objectif, et ces deux sens pourraient ou non pointer vers une même référence. D'où une seconde remarque, qui a

trait à la signification du verbe « être » (ou « être identique à ») dans le contexte de ce problème. Il ne s'agit pas forcément de pointer, à travers deux sens différents, vers une même référence (*je* ne suis numériquement identique à *cette* personne que si et seulement si une position moniste sur la personne est correcte). Mais, il s'agit tout au moins de pointer vers deux références en association relativement constante mais qui pourraient être disjointes (dans une conception dualiste de la personne, *je* ne suis pas numériquement identique à *cette* personne, mais *je* suis néanmoins relié d'une manière ou d'une autre à *cette* personne). Enfin, une troisième remarque a trait à la signification des mots « moi » et « je ». Ceux-ci renvoient nécessairement à l'énonciateur des propositions qui les contiennent (ce sont des concepts dits « indexicaux »). Si Paul dit : « Je suis triste », toute personne comprendra que c'est Paul, c'est-à-dire la personne qui vient de prononcer cette phrase, pas une autre, qui est triste. Sur cette base linguistique, on pourrait vouloir également contester la pertinence du problème, comme illusoire, et consistant en une réification abusive de termes du langage naturel. Ainsi, si je me demande : « Comment se fait-il que je sois cette personne ? », on pourrait vouloir répondre à cette question de la façon suivante : Cette personne énonce une proposition, et le mot « je » qu'elle utilise dans sa proposition renvoie nécessairement à elle-même, par définition. Ainsi, on peut remplacer « je » par « cette personne » dans la proposition, ce qui donne : « Comment se fait-il que cette personne est cette personne ? ». Puisque, d'après le principe d'identité, A=A pour tout A, alors « cette personne est cette personne » est une nécessité logique. Donc, « je suis cette personne » est une nécessité logique. Ce serait néanmoins confondre le point de vue en 1ère personne et celui en 3ème personne. Lorsque je dis « je » ou « moi », je renvoie à moi-même en tant que sujet. Lorsque je dis « cette personne » en me désignant, je renvoie à moi-même en tant qu'objet public. Ces deux aspects, nous l'avons dit plus haut, ne sont pas nécessairement liés. Nous pourrions concevoir, à travers l'expérience de pensée de la désincarnation

progressive, d'être un pur sujet auquel ne correspond aucun objet public situé dans l'espace-temps.

Imaginez le scénario suivant : Vous vous réveillez, un matin, et en vous regardant dans le miroir pour vous brosser les dents, vous vous rendez compte que votre reflet n'y apparaît pas ! A la place, vous n'y voyez que le meuble et le mur, situés en face, ainsi que le reflet de votre brosse à dents, qui semble flotter dans les airs. Affolé, vous poussez un cri qui réveille votre compagne/compagnon. Celle/Celui-ci regarde dans votre direction mais ne semble pas vous voir, puis croit rêver en voyant votre brosse à dents semblant être, pour elle/lui, en suspension dans les airs. Puis, soudain, celle-ci traverse vos mains et tombe dans le lavabo. Vous essayez de la reprendre, mais, à votre grande surprise, c'est impossible, vos mains semblent passer au travers, sans résistance, comme à travers un hologramme. Puis, vous vous rendez compte que vos mains, ainsi que tout le restant de votre corps, s'effacent petit à petit. Au bout de quelques secondes, vous êtes devenu entièrement invisible. Vous appelez votre compagne/compagnon, mais, cette fois-ci, bizarrement, il/elle ne semble pas vous avoir entendu. Réitérant votre appel à plusieurs reprises, vous avez la sensation que votre voix se fait de moins en moins perceptible, jusqu'à l'extinction totale ! De plus en plus paniqué, vous vous dirigez vers la porte de votre appartement pour sortir. Ce faisant, vous avez la curieuse sensation de devenir de plus en plus léger, le sol sous vos pieds – invisibles – se fait de moins en moins sentir. Au bout de quelques secondes, vous vous sentez tel un astronaute dans l'espace, flottant sans plus ressentir le moindre effet de la gravité, une simple poussée de vos « jambes » ou de vos « bras » vous propulse en ligne droite. Arrivé à la porte de votre appartement, terrifié, vous vous rendez compte que vous ne pouvez pas saisir la clenche. Qu'à cela ne tienne, vous traversez sans peine la porte qui, pour vous, ne semble plus être qu'un hologramme, et vous vous retrouvez dans la rue, où

les gens, les vélos, les voitures, etc. vous traversent sans heurt, et sans remarquer votre présence. Non seulement vous êtes à présent complètement invisible mais vous êtes devenu intangible, votre « corps » semble traverser sans peine tout matériau. La contrepartie est que vous ne pouvez plus du tout agir sur quoi ou qui que ce soit.

Il semble donc que l'on peut concevoir de manière logiquement consistante, c'est-à-dire sans contradiction (du moins, pas de prime abord), une disjonction entre les versants subjectifs et objectifs de la personne. Cela n'implique en rien une possibilité physique effective.

Cela n'implique pas, non plus, que l'on aurait pu être quelqu'un d'autre, car être quelqu'un d'autre implique, normalement, d'avoir d'autres souvenirs. De ce fait, on a du mal à isoler un « moi » quintessenciel qui pourrait donner un sens à cette expression.

On croit pouvoir s'imaginer que l'on aurait pu être quelqu'un d'autre et que quelqu'un d'autre aurait pu être nous. De cette croyance en découle une autre selon laquelle nous sommes essentiellement un « point de vue » incarné, un substrat indéterminé, un « spectateur passif » qui contemple les processus cérébraux impersonnels d'un certain individu, les transmuant en processus conscients personnels, mais qui aurait pu être associé à un autre corps et contempler les processus cérébraux d'un autre individu. Nous appellerons **« moi » métaphysique** ou **transcendantal** le « moi » ainsi défini.

On peut contraster cette conception abstraite du « moi » avec une autre conception, plus proche du sens commun, qui fait du « moi » une entité concrète, caractérisée par un ensemble de caractéristiques objectives qui évolue dans une certaine continuité au cours du

temps. Nous appellerons **« moi » biopsychologique** ou **empirique** ce deuxième type de « moi ».

En réalité, lorsque l'on croit s'imaginer être quelqu'un d'autre, on se l'imagine toujours de notre point de vue. On ne peut supprimer, pendant que l'on tente d'imaginer que l'on est quelqu'un d'autre, les souvenirs et traits psychologiques de la personne que l'on est vraiment, ainsi, tout simplement, que son corps.

On peut, semble-t-il, s'imaginer une **transformation** en quelqu'un d'autre. Imaginons une personne A et une personne B. Supposons que l'on déclenche un processus qui permette de transformer très progressivement et simultanément la structure moléculaire de A pour le changer en un individu en tout point identique à B ; et la structure moléculaire de B pour le changer en un individu en tout point identique à A. Appelons AB l'individu A en cours de transformation vers un individu en tout point identique à B et BA l'individu B en cours de transformation vers un individu en tout point identique à A. Au cours du processus, on supposera que non seulement l'apparence physique mais également les souvenirs et la personnalité sont modifiés. Au final, on a donc un résultat impeccablement parfait : AB est convaincu d'être B et BA est convaincu d'être A. Pour autant, lorsque la transformation commence, AB s'identifie sans doute et fermement à A en train de subir des modifications en lui-même qui le font ressembler à B et ne s'identifie sans doute pas du tout à B. Il en est de même pour BA en un sens inverse : au début il s'identifie à B en train de subir une aliénation de sa personne, et non pas à A.

La question qui nous intéressera est la suivante : à partir de quel moment l'identification de AB et de BA à, respectivement, A et B, s'intervertit-elle ? Il n'est pas

évident de répondre à cette question, mais il y a sans doute une réponse, un moment où l'interversion a lieu. Ce moment n'est d'ailleurs pas forcément un instant. Cela peut être une certaine période de temps étalée durant laquelle règne, chez AB et chez BA, une grande confusion mentale concernant qui ils sont vraiment. On peut supposer que cela commence, chez AB, lorsque les souvenirs de A sont remplacés par les souvenirs de B, et, symétriquement chez BA, lorsque les souvenirs de B sont remplacés par les souvenirs de A. Ce qui va être particulièrement décisif, cela va être l'interversion des souvenirs du **processus lui-même de transformation**.

A partir du moment où AB ne se souviendra plus d'être en voie de transformation de A vers B et aura le souvenir d'être en voie de transformation de B vers A, il s'identifiera vraisemblablement à B. De même, lorsque le souvenir de BA d'être en train de se transformer de B vers A aura été remplacé par le souvenir d'une transformation de A vers B, BA s'identifiera certainement à A.

A moins bien sûr que AB et BA soient un peu philosophes, et notamment des philosophes antiréductionnistes. Si AB et BA sont des philosophes, ils auront peut-être anticipé cette curieuse impression d'interversion des « moi » lorsqu'un certain niveau de transformation est atteint. S'ils sont antiréductionnistes, ils considèreront que cette impression d'interversion des « moi » est illusoire et concluront, contre les apparences,

qu'ils sont A s'ils ressemblent en tout point à B ou qu'ils sont B s'ils ressemblent en tout point à A.[72]

Il n'y a, cependant, aucun moyen de vérifier pour départager ces hypothèses. Seule la connaissance du processus de transformation qui a eu lieu – donc une connaissance de nature historique – peut apporter de l'eau au moulin de l'interprétation antiréductionniste[73]. Mais empiriquement, tout porte à croire que AB, en fin de parcours est B, et que BA, en fin de parcours est A, et donc qu'il y a bel et bien eu interversion des « moi ». En fait, même si AB est réellement A et que BA est réellement B, tout se passe comme si AB était B et BA était A.

L'antiréductionniste qui persisterait à croire que, dans notre exemple, le « moi » essentiel de A se retrouve à présent en B et que le « moi » essentiel de B se retrouve à présent en A doit faire face à un autre problème de taille. En effet, il est contraint d'admettre que tout ce qui fait d'une personne (au sens ordinaire du terme) la personne qu'elle est, c'est-à-dire, par exemple, quelqu'un d'intelligent, de bon, apprécié par ses amis, ses parents, etc. n'est pas elle-même, ne fait pas partie de ce qu'elle est essentiellement.

---

72 Il faut signaler que des gens pensant que notre identité personnelle s'arrime sur la fraction de matière qui nous constitue, ainsi que sur les fractions de matière en continuité – par renouvellement progressif au sein de l'organisme – de cette fraction de matière, peuvent, sans être à proprement parler antiréductionnistes (de prime abord), aboutir à cette même conclusion sur l'identité de A et de B. Il en est de même pour les gens qui conçoivent l'identité personnelle comme reposant sur une certaine continuité, au cours du temps, de l'évolution de certaines caractéristiques. L'antiréductionnisme ne semble donc pas nécessaire pour donner un sens à l'expression « être quelqu'un d'autre », du moment qu'on la comprend comme dans cette expérience de pensée, c'est-à-dire comme : « devenir quelqu'un d'autre » et non pas « avoir pu être quelqu'un d'autre ».

73 Ainsi qu'aux interprétations basées sur les conceptions évoquées dans la note précédente.

Bref, nous ne sommes pas essentiellement notre « moi » biopsychologique ou empirique. Nous sommes essentiellement notre « moi » métaphysique ou transcendantal.

Si l'antiréductionnisme est vrai, les caractéristiques objectives d'une personne (au sens ordinaire du terme) sont, au mieux, un indice de l'identité personnelle, mais même pas une preuve. Il serait possible que, par exemple, les « moi » métaphysiques d'Oussama Ben Laden et de l'abbé Pierre permutent, sans que quiconque, pas même eux, ne puissent s'en rendre compte. D'ailleurs, les « moi » métaphysiques des 6 milliards d'êtres humains de la Terre pourraient très bien permuter ainsi toutes les 15 secondes, sans que cela change quoi que ce soit concrètement[74]. Leibniz l'avait très bien vu qui écrit, à travers le personnage de Théophile, dans les *Nouveaux Essais sur l'Entendement Humain*[75] :

> « [I]l se peut que dans un autre lieu de l'univers ou dans un autre temps, il se trouve un globe qui ne diffère point sensiblement de ce globe de la terre où nous habitons, et que chacun des hommes qui l'habitent ne diffère point sensiblement de chacun de nous qui lui répond. Ainsi il y aura à la fois plus de cent millions de paires de personnes semblables, c'est-à-dire de deux personnes avec les mêmes apparences et consciences ; et Dieu pourrait transférer les esprits seuls ou avec leur corps d'un globe dans l'autre sans qu'ils s'en aperçussent »

C'est complètement contre-intuitif. En pratique, on juge qu'une personne est la personne qu'elle est sur la base de caractéristiques objectives. C'est une personne concrète,

---

[74] Les conceptions de l'identité personnelle évoquées dans les deux notes précédentes posent le même problème : une fraction de matière indifférenciée ne semble guère être un meilleur candidat à nos intuitions sur ce qui est essentiel en nous-même (sans compter que cette fraction de matière se renouvelle au cours du temps), qu'un « moi » métaphysique évanescent. Toutefois, il apparaît plus difficile d'envisager une désolidarisation entre la matière de notre corps et les caractéristiques objectives qu'elle supporte, qu'entre ces dernières et notre « moi » métaphysique hypothétique.

[75] LEIBNIZ, Gottfried Wilhelm, *Nouveaux Essais sur l'Entendement Humain*, Livre II, Chapitre XXVII, p. 190

un « moi » psychologique – c'est-à-dire un ensemble dynamique de caractéristiques objectives – et non pas métaphysique, que l'on va, par exemple, aimer, détester, percevoir, désirer, envier, manipuler, flatter, etc.

On en vient à la célèbre remarque de Blaise Pascal[76] :

> « Qu'est-ce que le moi ?
>
> Un homme qui se met à la fenêtre pour voir les passants, si je passe par là, puis-je dire qu'il s'est mis là pour me voir ? Non ; car il ne pense pas à moi en particulier. Mais celui qui aime quelqu'un à cause de sa beauté, l'aime-t-il ? Non; car la petite vérole, qui tuera la beauté sans tuer la personne, fera qu'il ne l'aimera plus.
>
> Et si on m'aime pour mon jugement, pour ma mémoire, m'aime-t-on, moi ? Non ; car je puis perdre ces qualités sans me perdre moi-même. Où est donc ce moi, s'il n'est ni dans le corps, ni dans l'âme ? Et comment aimer le corps ou l'âme, sinon pour ces qualités qui ne sont point ce qui fait le moi, puisqu'elles sont périssables ? Car aimerait-on la substance de l'âme d'une personne abstraitement, et quelques qualités qui y fussent ? Cela ne se peut, et serait injuste. On n'aime donc jamais personne, mais seulement des qualités.
>
> Qu'on ne se moque donc plus de ceux qui se font honorer pour des charges et des offices, car on n'aime personne que pour des qualités empruntées. »

Si on récapitule ces deux conceptions du « moi », on peut distinguer :

- Un « moi » biopsychologique ou empirique : c'est une continuité temporelle de caractéristiques objectives ;

---

76 PASCAL, Blaise. *Pensées*.

- Un « moi » métaphysique ou transcendantal : c'est ce qui reste du précédent, lorsque l'on lui a ôté toutes les caractéristiques objectives.

Un autre problème surgit. On oublie souvent que les localisations spatiales et temporelles sont des caractéristiques objectives également (quoique extrinsèques). Le « moi », tel que le conçoit l'antiréductionniste est-il atemporel et aspatial ? Il faut dire que s'il ne présente aucune caractéristique objective intrinsèque, on comprend mal comment il pourrait être situé quelque part dans l'espace et le temps[77]. Mais si le « moi » métaphysique est atemporel et aspatial, alors il semble qu'il est aussi, par le fait, unique et commun à tout le monde. Dès lors, on ne comprend plus comment il peut fonder notre identité personnelle, comment il peut se distinguer d'une personne à l'autre.

Peut-être qu'il y a un sens à parler d'un « moi » métaphysique, comme il y a un sens à parler d'un ensemble vide. Mais il semble alors que, tel l'ensemble vide qui est à la fois unique et élément de tous les ensembles, le « moi » métaphysique est unique et élément de tous les « moi ». A moins de lui supposer quelques qualités « occultes », bien entendu...

Et en l'occurrence, entre le concept de « moi » biopsychologique et celui de « moi » métaphysique, il y a peut-être un sens à parler d'un autre type de « moi » qui pourrait – ou aurait pu – être différentes personnes.

Dans l'expérience de pensée de la désincarnation progressive, nous avons suggéré l'existence d'un type de « moi » particulier. Il ne se réduit pas à une sorte de soubassement

---

[77] Les conceptions de l'identité personnelle évoquées dans les trois notes précédentes ne rencontrent évidemment pas ce problème.

commun à toutes les personnes sans être personne en particulier, tel le « moi » métaphysique, mais il ne correspond pas non plus à des caractéristiques objectives, tel le « moi » physique. Ce « moi » a la propriété de percevoir mais de ne pas être perçu. Il a donc du contenu, mais ce contenu est irréductiblement privé. Nous pouvons proposer à présent de distinguer entre :

- Un « moi » biopsychologique ou empirique : c'est une continuité temporelle de caractéristiques objectives et subjectives ;

- Un « moi » phénoménologique ou semi-métaphysique ou semi-transcendantal : c'est ce qui reste du précédent, lorsque l'on a ôté toutes les caractéristiques objectives, mais que l'on a gardé les caractéristiques subjectives.[78]

- Un « moi » métaphysique proprement dit, ou hyper-métaphysique ou hyper-transcendantal : c'est ce qui reste du précédent, lorsque l'on a ôté toutes les caractéristiques subjectives.[79]

Compliquons un peu notre scénario de départ : cela fait à présent un bon bout de temps que vous vous trouvez dans votre nouvel état. Avec un certain entraînement, vous

---

[78] On pourrait imaginer aussi, bien entendu, un « moi » obtenu par soustraction des seules caractéristiques subjectives, laissant les caractéristiques objectives intactes. Nous pouvons appeler ce type de « moi » : « moi » zombie.

[79] On pourrait ajouter également, à ce tableau, un « moi » correspondant à la fraction de matière qui nous constitue, ainsi qu'aux fractions de matière en continuité – par renouvellement progressif au sein de l'organisme – de cette fraction de matière. Nous pouvons appeler ce type de « moi » : « moi » physique ou matériel.

avez remarqué que vous arriviez à vous « identifier » plus ou moins complètement, pendant quelques minutes, aux personnes que vous traversiez. Au début, vous captez, pendant quelques secondes, quelques-unes de leurs pensées, sensations et émotions, quelques-uns de leurs souvenirs, puis vous arrivez plus complètement à être eux, faisant passer vos propres souvenirs, pensées, émotions, au second plan, puis, les faisant disparaître pendant quelques petites minutes. Lorsque vous ressortez des gens, vous reprenez petit à petit conscience de vous-mêmes, chargé de toute une vie en supplément. Grâce à cette capacité, vous avez ainsi réussi à devenir, pendant quelques instants, un paysan auvergnat, un informaticien chez Apple, une femme enceinte, un aborigène australien, le roi de Belgique, Margaret Thatcher, etc. Vous pourriez à présent écrire des livres entiers sur la vie de ces gens-là, leurs secrets intimes, leurs désirs profonds, leurs pensées cachées, etc.

Lorsque vous vous incarnez dans tous ces personnages, vous vous oubliez vous-mêmes complètement. Cependant, lorsque vous vous désincarnez, vous recouvrez progressivement votre mémoire, sans toutefois oublier ce que vous avez vécu lors de vos incarnations[80].

Ces souvenirs des vies que vous traversez sont authentiques ou non. S'ils sont authentiques, comme il semble bien que ce soit le cas, d'après le scénario évoqué, alors il semble bien qu'il y a un sens à dire que, pendant vos incarnations, vous parvenez à être quelqu'un d'autre que la personne que vous êtes, et que vous y parvenez pleinement,

[80] Sans chercher bien loin, on observe ce genre de phénomène curieux avec les rêves. Lorsque l'on rêve, on n'est rarement conscient d'être en train de rêver, et donc de dormir. Qui plus est, les liens avec les journées passées ne sont pas toujours franchement évidents. En revanche, une fois éveillé, ces liens se rétablissent. Qui plus est, on est à présent sûr d'être éveillé, et il est assez fréquent, au moins les premières minutes, de se rappeler de nos rêves de la nuit. Le sujet du rêve et le sujet de la veille sont sans doute différents sur bien des points. Peut-on douter, cependant, de leur identité numérique ?

puisque, durant ces périodes de temps, vous en oubliez jusqu'au fait que vous êtes en train de vivre la vie de quelqu'un d'autre. Donc, pendant ces moments-là, vous êtes le quelqu'un d'autre en question. S'ils ne sont pas authentiques, alors qu'est-ce qu'un souvenir authentique ?[81]

## *E. Intuition n°5 : Les désignateurs rigides.*

*« Au fond, peu importe quelle est la nature du « moi » métaphysique, c'est l'entité, quelle qu'elle soit, qui est désignée par mon nom propre (au sens kripkéen), et ce nom propre a une signification bien précise et stable dans le temps, et à travers les mondes possibles. »*

Critique :

L'existence d'un nom propre n'implique pas l'existence d'un sens (*a fortiori* d'un référent) stable et précis à ce nom propre, et un « moi » métaphysique disposant de qualités occultes propre à le différencier d'autres « moi » métaphysiques a de bonnes chances de n'être qu'une illusion linguistique et à n'avoir qu'une existence purement verbale.

[81] Dans les conceptions de l'identité personnelle qui ne reposent que sur une certaine continuité au cours du temps dans l'évolution de certaines caractéristiques, il est possible de concevoir, non pas avoir été quelqu'un d'autre, mais devenir quelqu'un d'autre, suite à une transformation progressive (cf. note 72). Dans les conceptions de l'identité personnelle reposant sur la fraction de matière qui nous constitue, ainsi que sur les fractions de matière en continuité – par renouvellement progressif au sein de l'organisme – de cette fraction de matière, il est possible de concevoir avoir été quelqu'un d'autre, si (l'essentiel de) la fraction de matière qui avait constitué le zygote que nous avons été, s'était trouvée à former le zygote ayant donné naissance à quelqu'un d'autre (par exemple à Albert Einstein). L'antiréductionnisme n'est donc, encore une fois, pas nécessaire pour donner un sens à ces intuitions.

Lorsque, avec un raisonnement valide et des prémisses correctes, on aboutit à une conclusion paradoxale, il faut essayer de s'interroger sur le sens des concepts invoqués dans les propositions.

Ce qu'il faut comprendre, c'est que ce qui est pratique dans une conversation ordinaire, peut être source de confusion si on veut raisonner logiquement. Pourquoi les sciences inventent-elles des nouveaux mots ? Parce qu'elles en ont besoin pour s'exprimer clairement. En utilisant les seules ressources du langage naturel, leurs descriptions seraient vagues, tortueuses, indigestes...

Un exemple de raisonnement incorrect permis par le langage naturel est connu sous le nom de « paradoxe du tas ». C'est un exemple de paradoxes dits « paradoxes sorites ». Si on admet que ce raisonnement est logiquement exact et que toutes les prémisses sont vraies mais que la conclusion est quand même fausse, alors où est l'erreur ? Dans le langage lui-même, qui est trop vague.

Voici l'exemple :

(1) un ensemble comportant 100000 grains de sable est un tas (prémisse de base)

(2) si un ensemble comportant n grains de sable est un tas, alors un ensemble comportant n - 1 grains de sable est un tas (prémisse d'induction)

(3) si un ensemble comportant 100000 grains de sable est un tas, alors un ensemble comportant 99999 grains de sable est un tas (de (2))

(4) \ un ensemble comportant 99999 grains de sable est un tas (de (1) et (3))

(5) si un ensemble comportant 99999 grains de sable est un tas, alors un ensemble comportant 99998 grains de sable est un tas (de (2))

(6) \ un ensemble comportant 99998 grains de sable est un tas (de (4) et (5))

(7) si un ensemble comportant 99998 grains de sable est un tas, alors un ensemble comportant 99997 grains de sable est un tas (de (2))

(8) \ un ensemble comportant 99997 grains de sable est un tas (de (6) et (7))

(9) ...

(10) \ un ensemble comportant 1 grain de sable est un tas

Les paradoxes des expériences de pensée sur l'identité personnelle ont une structure assez similaire, sauf que cela ne se voit pas forcément du premier coup d'oeil, car autant il est intuitif de considérer la notion de « tas » comme imprécise ou reflétant une réalité imprécise, autant il semble contre-intuitif de penser que la notion de « moi » n'ait pas un sens précis non plus, ou ne renvoie pas à une réalité précise.

On peut penser que c'est parce que, dans des situations normales de la vie de tous les jours, ce sens est effectivement précis, il désigne le noyau relativement permanent de l'individu concret, dont la localisation spatiale varie en continu, que l'on peut reconnaître et qui se reconnaît lui-même, l'ensemble des souvenirs, désirs, idées qui lui sont propres et qui

changent lentement. De plus, le « moi » ainsi défini est source et objet d'attachement, d'amour, de haine, qui renforcent ce sentiment d'ipséité, d'absolue singularité irremplaçable. Si un être cher à vos yeux perdait son « moi » métaphysique, en acquérait un autre (disons celui d'un malfrat), et qu'objectivement il n'y ait aucun changement décelable, ni pour vous, ni pour lui ni pour quiconque, cesseriez-vous de l'aimer ?

Les expériences de pensée sur l'identité personnelle rendent ce sens inopérant, mais alors que dans le cas du tas de sable qui ne comporte plus aucun grain, on voit que c'est la notion de « tas » qui est en cause, dans le cas du « moi » qui n'a plus une seule de ses caractéristiques de départ, on ne veut pas croire que ce soit le concept de « moi » qui soit en cause, et on préfère imaginer qu'il y a un sens à parler d'un « moi » qui n'est plus personne ou plus la personne qu'il désignait originellement, et on invente donc une entité éthérée qui serait le siège permanent de ce « moi » qui n'est personne.

### *F. Intuition n°6 : Les approches en 1ère et 3ème personne sont irréductibles. Chaque approche en 1ère personne est irréductible à une autre approche en 1ère personne.*

*« La différence entre ma conscience et mon cerveau semble incommensurable. Ainsi, je suis sûr de l'existence de ma conscience et du fait que c'est ma conscience, alors que je peux douter de l'existence de mon cerveau et du fait qu'il s'agit bien de mon cerveau.*

*Ensuite, la différence entre mon cerveau et celui de mon double microphysique est faible, sinon nulle. De mon point de vue, ils m'apparaissent être des cerveaux indiscernables. Je peux, par exemple, percevoir*

*une configuration particulière dans mon cerveau, de la même façon que je peux la percevoir sur celui de mon double microphysique.*

*En revanche, si j'ai de bonnes raisons de penser que nos deux cerveaux existent, et si je suis certain de l'existence de ma conscience, je suis encore moins sûr de l'existence de la conscience de mon double microphysique que de l'existence de nos cerveaux. Je ne peux ressentir une éventuelle expérience de douleur à ses dents présente dans sa conscience, telle que lui la ressent, et elle m'importe peu.*

*Il y a donc un clivage profond entre, d'une part, le domaine matériel des cerveaux, corps, etc. et celui, spirituel, des consciences, qui justifie un certain dualisme épistémologique sinon ontologique, et d'autre part, chaque conscience par rapport aux autres consciences, qui justifie un pluralisme épistémologique voire ontologique des consciences. Autrement dit, une âme est un type de chose différente d'un corps, mais encore, chaque âme est un type de chose différent d'une autre âme, si bien que la question de l'identité personnelle se pose : pourquoi telle âme habite-t-elle tel corps et pas un autre ? »*

Critique :

Ce type de raisonnement naît d'une confusion que l'on pourrait qualifier de : confusion entre la carte et le territoire.

Considérons le cas de 3 individus a, b, c qui sont des triplés microphysiques.

Il y a 3 territoires : a, b et c dans leur ipséité respective, leur nature profonde et ontologique si vous préférez.

A ces 3 territoires correspondent 18 cartes. Soit un individu x, un individu y, et un mode d'appréhension (subjectif ou objectif), ces cartes correspondent à : comment x apparaît subjectivement et objectivement à y. Par mesure de simplification et de clarté, nous utiliserons l'écriture suivante pour énumérer toutes les cartes :

xSy = Comment x apparaît subjectivement à y

xOy = Comment x apparaît objectivement à y

Par « apparaître objectivement », il faut entendre une description « de l'extérieur » d'un sujet donné, c'est-à-dire une description de ce que l'on perçoit (par les sens) de ce sujet : est-il grand, petit, chauve, chevelu, aimable, antipathique, etc. Par « apparaître subjectivement », il faut entendre une description « de l'intérieur » d'un sujet donné, c'est-à-dire une description de ce qu'il perçoit (par les sens). Voit-il un arbre à la gauche de son champ de vision, perçoit-il le bleu du ciel, expérimente-t-il une douleur dans sa jambe, etc. ? Il s'agit donc de se mettre à sa place, pour saisir le « what it is like to be » cher à Thomas Nagel.

Les 18 cartes sont les suivantes :

| | | | |
|---|---|---|---|
| 1) aSa | 3) cSc | 5) bOb | 7) aSb |
| 2) bSb | 4) aOa | 6) cOc | 8) aSc |

| | | | |
|---|---|---|---|
| 9) bSa | 12) cSb | 15) bOa | 18) cOc |
| 10) bSc | 13) aOb | 16) bOc | |
| 11) cSa | 14) aOc | 17) cOa | |

Si maintenant on compare toutes ces cartes, on peut les regrouper par groupes de ressemblance :

1 – 2 – 3 = descriptions auto-subjectives

4 – 5 – 6 = descriptions auto-objectives

7 – 8 – 9 – 10 – 11 – 12 = descriptions hétéro-subjectives

13 – 14 – 15 – 16 – 17 – 18 = descriptions hétéro-objectives

A supposer que a, b et c ne soient pas des partisans de la survenance du mental sur le physique (à toute différence mentale correspond une différence physique) et qu'ils ne soient pas particulièrement empathiques, ils ne pourront pas dire grand chose sur comment leurs alter ego s'apparaissent subjectivement à eux-mêmes. S'ils ne disposent pas de miroir ou d'appareils photos, ils auront également plus de mal à se décrire eux-mêmes objectivement qu'à décrire les deux autres. On a donc bien, à peu près, ce regroupement. En pratique, il est cependant plus facile d'avoir une bonne description auto-objective qui se rapproche d'une description hétéro-objective que d'avoir une bonne description hétéro-subjective qui se rapproche d'une description auto-subjective.

On peut donc regrouper nos descriptions autrement :

1 – 2 – 3 = descriptions auto-subjectives

7 – 8 – 9 – 10 – 11 – 12 = descriptions hétéro-subjectives

4 – 5 – 6 – 13 – 14 – 15 – 16 – 17 – 18 = descriptions objectives

L'impression de dualisme et d'irréductibilité de l'identité personnelle vient de ce que l'on a le sentiment qu'il existe quelque chose de public, qui est décrit de la même manière par tous (4 – 5 – 6 – 13 – 14 – 15 – 16 – 17 – 18), et qu'il existe quelque chose de privé, auquel seule une personne à accès directement (1 – 2 – 3 et 7 – 8 – 9 – 10 – 11 – 12), les autres personnes ne pouvant que l'inférer à partir de la chose publique.

Cette impression, selon nous, est trompeuse. En effet, nous avons supposé implicitement que a, b et c n'étaient pas particulièrement empathiques, mais on pourrait imaginer qu'ils sont dotés d'une empathie particulièrement aiguë. Si c'est le cas, alors ils sauraient exactement comment leurs alter ego s'apparaissent à eux-mêmes subjectivement. Notez qu'il ne s'agira pas alors d'une inférence. Une personne empathique ne se dit pas : « il pleure donc il est triste donc je suis triste ». La vision de larmes sur le visage d'autrui produira directement de la tristesse chez lui, sans qu'il ait besoin de raisonner. C'est direct et automatique. Parfois même, c'est subconsciemment que cela se joue et des personnes pourront se sentir triste en compagnie de certaines autres personnes, sans comprendre pourquoi. Vérification faite, ces personnes étaient tristes elles-mêmes mais ça n'apparaissait que très subtilement dans leur comportement. On peut penser que ce genre de chose est

très fréquent et est responsable de nombreuses fausses croyances d'avoir échangé ce sentiment par télépathie (nous n'affirmons pas que la télépathie n'existe pas dans certains cas, mais néanmoins, dans ces cas là, l'hypothèse d'une empathie inconsciente est de loin la plus plausible).

Sous l'hypothèse d'une parfaite empathie, on aurait donc le regroupement suivant :

1 – 2 – 3 – 7 – 8 – 9 – 10 – 11 – 12 = descriptions subjectives

4 – 5 – 6 – 13 – 14 – 15 – 16 – 17 – 18 = descriptions objectives

Il reste que l'on peut encore se demander pourquoi on a encore deux séries de descriptions qui ne coïncident pas. On peut penser que c'est toujours encore une fois le même problème de carte et de territoire. Imaginons que a, b et c aient à décrire une maison, sans avoir le même point de vue sur elle. L'individu « a » la voit de face et disons à 10 m. L'individu « b » la voit de côté et à 100 m. L'individu « c » la voit de dessus et à 15 m. Si on compare les croquis, on pourra très bien penser que ce sont des croquis de trois maisons différentes, alors qu'en fait, ce sont des croquis d'une seule maison. Il en est de même entre la série de descriptions 1 – 2 – 3 – 7 – 8 – 9 – 10 – 11 – 12 et la série de descriptions 4 – 5 – 6 – 13 – 14 – 15 – 16 – 17 – 18, elles ne sont pas faites du même point de vue. Dans la première série, il s'agit de descriptions « de l'intérieur », c'est-à-dire de descriptions de l'effet que cela fait d'*être* une personne donnée, tandis que dans la seconde série, il s'agit de descriptions « de l'extérieur », c'est-à-dire de descriptions de l'effet que cela fait de *percevoir* une personne donnée.

Notez que l'empathie peut être vue comme un sens qui permet d'être littéralement, quoique en partie, une autre personne. On répondra peut-être : « Oui, mais quand on ressent la souffrance de quelqu'un d'autre, c'est toujours de notre point de vue qu'on la ressent ». Bien sûr, mais lorsque l'on perçoit quelqu'un d'autre, c'est également de notre point de vue que nous le percevons.

## *G. Intuition n°7 : Etre une PERSONNE, c'est être UNE personne.*

*« Les individus A11, A12, A13, A14 et A15 ne peuvent pas tous être A1 simultanément, car cela voudrait dire qu'A1 est à présent une créature curieuse avec 5 têtes, 10 yeux, 10 bras, etc. ce qui est absurde car A11, A12, A13, A14 et A15 ont chacun le sentiment d'être différent des 4 autres, ils peuvent mener des existences complètement séparées sans interagir, et ils apparaissent aux yeux de tous comme 5 personnes bien différentes, et non comme une personne dans 5 corps et à 5 endroits différents. Pour que A1 soit à la fois A11, A12, A13, A14 et A15, il faudrait qu'il puisse se sentir vivre à travers ces 5 individus différents, qu'il se sente exister à 5 endroits à la fois. Pour que cela soit possible, il faudrait que ces 5 individus soient constamment connectés par une sorte de relation télépathique. »*

Critique :

Nul besoin de supposer un lien télépathique entre individus pour donner sens au fait que plusieurs individus puissent être une seule personne.

Pour comprendre cela, nous pouvons imaginer une petite expérience de pensée.

Nous nous placerons, pour commencer, dans un cadre de pensée dualiste. Nous justifierons ce choix ultérieurement. Dans ce cadre dualiste, chaque individu a une âme. Supposons toutefois une petite entorse, une anomalie. Quelque part dans le monde se trouve un individu (nous l'appellerons Michel) qui, depuis sa naissance, est doté de deux âmes, numériquement différentes. Appelons-les « âme A » et « âme B ».

Chacune de ces âmes croit être « seule à bord » car il se trouve qu'elles sont qualitativement identiques (étant influencées dès le départ par exactement les mêmes gènes et le même environnement), associées au même corps et que, pour ainsi dire, elles se superposent spatialement (pour autant que cela puisse vouloir dire quelque chose pour un être immatériel d'occuper une position dans l'espace, et *a fortiori* la même position dans l'espace qu'un autre être immatériel). Nous négligerons l'existence possible d'un libre-arbitre qui pourrait les arracher à ce déterminisme biosocial.

Lorsque Michel dit à quelqu'un : « Pff... **je** suis envahi d'un profond spleen en cette grise soirée d'automne où les feuilles mortes viennent tomber nonchalamment sur le sol imbibé d'eau de pluie », à qui ce **je** réfère-t-il ? A l'âme A ? A l'âme B ? Aux deux âmes à la fois ? A aucune des deux ?

Objectivement aux deux, puisque ce sont les deux qui sont envahies d'un profond spleen (elles sont qualitativement identiques, rappelons-le) et, si une seule des deux existait dans le corps de Michel, cela aurait suffit pour qu'il exprime exactement le même sentiment de spleen de la même façon. Pourtant, chacune des deux âmes pensera que ce **je** se réfère à elle et à elle seule puisqu'elle ignore l'existence de l'autre âme.

Maintenant, imaginons que par une opération du Saint Esprit, le corps de Michel se trouve coupé en deux et que chaque partie reconstitue la partie manquante à l'identique. Imaginons que durant cette opération, A et B se trouvent séparés. A est associée avec l'individu formé avec la partie droite du corps du Michel original, tandis que B est associée avec l'individu formé avec la partie gauche du corps du Michel original. A présent on a un Michel avec l'âme A (Michel-A) et un Michel avec l'âme B (Michel-B). Cette identité qualitative et de localisation spatiale étant rompue, chaque âme peut, à sa grande surprise peut-être, découvrir l'existence de son alter ego, ou se poser d'angoissantes questions métaphysiques sur l'identité personnelle, puisque l'âme A se rendra bien compte que l'âme qui anime Michel-B ne peut pas être elle, et l'âme B, de son côté, fera un constat similaire.

Maintenant, posons-nous la question suivante :

Qui, de Michel-A et de Michel-B est le vrai Michel ? Pour répondre à cette question, encore faudrait-il pouvoir définir ce que c'est d'être le vrai Michel. Si l'on dit que le vrai Michel c'est ce Michel qui était envahi par le spleen en cette grise soirée d'automne, etc. alors il y a objectivement deux vrais Michel qui étaient confondus (âme A et âme B) et qui sont maintenant séparés, un peu comme deux frères siamois qui seraient tellement soudés et fusionnés qu'ils n'auraient finalement formés... qu'un seul individu tout à fait normal !

Si l'on dit que le vrai Michel c'est ce Michel qui est formé avec la partie droite (ou pourquoi pas gauche) de ce Michel qui était envahi par le spleen en cette grise soirée d'automne, etc. alors il n'y a qu'un seul vrai Michel, mais notre critère pour le désigner est totalement arbitraire (pourquoi telle partie du corps de Michel serait-elle plus importante qu'une autre au regard de son identité ?). Le critère précédent est donc préférable.

Supposons maintenant que l'on dise à Michel, avant qu'il soit découpé en deux, que l'individu formé avec la partie droite de son corps va être envoyé au Pôle Nord, tandis que l'individu formé avec la partie gauche de son corps va aller au Pôle Sud.

On demande ensuite à Michel : « Iras-**tu** au Pôle Nord ou Pôle Sud ? ». Là encore, la réponse à la question dépend de qui est désigné par **tu.** Objectivement, ce sont les deux âmes qui sont désignées par **tu** mais chacune pensera que l'on ne parle que d'elle et d'elle seule. Il est donc juste de dire que Michel ira à la fois au Pôle Sud et au Pôle Nord (en ce sens trivial que le Michel qui va au Pôle Nord est tout autant le vrai Michel que celui qui va au Pôle Sud), et en même temps, il est juste de dire que Michel ira seulement à un des deux endroits, mais qu'il ne sait pas lequel, car il (en tant qu'âme A ou B) ne sait pas qui il est, c'est-à-dire à quelle partie de son corps il sera associé. C'est seulement après l'opération qu'il le saura. Ainsi, le Michel formé avec la partie droite du corps du Michel original se dira : « **Je** suis Michel. **Je** suis au Pôle Nord, donc **je** suis formé avec la partie droite de **mon** ancien corps » (ici, **je** et **mon** désignent l'âme A) ; tandis que le Michel formé avec la partie gauche du Michel original se dira : « **Je** suis Michel. **Je** suis au Pôle Sud, donc **je** suis formé avec la partie gauche de **mon** ancien corps » (ici, **je** et **mon** désignent l'âme B). Mais ce que chaque Michel saura qu'il est, n'importe qui pourra, en principe, le savoir (**Michel** et **ancien corps**, dans les deux phrases, désignent à la fois l'âme A et l'âme B, c'est-à-dire le Michel concret et public). Ce n'est donc, en aucun cas, une connaissance privée.

Notre expérience de pensée considérait qu'il existait deux âmes en Michel, mais :

- Cela pourrait être le cas de tout le monde et cela ne changerait rien à ce que l'on peut empiriquement constater dans le monde ;

- Au lieu de deux âmes, il pourrait en exister plus, beaucoup plus, et même un nombre indéfini, et cela ne changerait rien à ce que l'on peut empiriquement constater dans le monde ;

- Enfin, toutes ces âmes pourraient avoir une ontologie naturaliste ; il n'est pas obligatoire qu'elles soient constituées d'une substance surnaturelle. Reprenons l'exemple des frères siamois. Supposons deux frères siamois (A et B) qui formeraient un organisme muni d'un corps et deux têtes. Il y a plusieurs manières de décrire ce corps : on peut dire que c'est le corps de A, que c'est le corps de B ou que c'est le corps de A et B. Toutes ces descriptions sont justes, même si elles n'ont pas le même sens, et il n'y a rien de magique à cela ! Donc, pour reprendre cette fois-ci notre expérience de pensée, il est possible de décrire le Michel avant duplication comme étant Michel-A, Michel-B ou Michel-A et -B. Toutes ces descriptions sont également justes, même si elles n'ont pas le même sens, et il n'y a rien de magique à cela non plus, si on y réfléchit bien. Il faut juste cesser de se représenter le temps comme une dimension foncièrement différente des dimensions de l'espace. Si on spatialise le temps, par exemple en traçant une ligne du temps sur une feuille de papier et en y représentant le Michel de départ et les deux Michel d'arrivée, le parallèle avec l'exemple des frères siamois saute aux yeux. Mais on peut trouver un exemple temporel aussi, si vous voulez, à la place de celui avec les frères siamois, et qui n'implique pas nécessairement de souscrire à une ontologie quadridimensionnelle de l'espace-temps. Si on considère par exemple la première année de médecine (PCEM1) en France, on peut la décrire comme la première année des

> études médicales ou comme la première année des études dentaires ou comme la première années des études médicales et dentaires. C'est, pour chaque étudiant réussissant le concours, son choix de carrière qui rétroactivement rendra, pour lui, une des descriptions plus pertinente que l'autre.

Ces dernières considérations nous suggèrent une nouvelle hypothèse suivant laquelle le problème de l'identité personnelle serait à la fois plus qu'un simple problème de mots et de convention linguistique, mais moins qu'un problème ontologique portant sur un référent réel de ces mots. Se serait un problème sémantique. A l'instar de l'étoile du matin et de l'étoile du soir, qui renvoient, toutes deux – mais selon deux sens différents – à la planète Vénus, il y aurait un sens objectif à parler de personne, et un sens subjectif. Ces deux sens ne renvoient pas à deux référents différents mais à deux façons d'appréhender un même référent.

Ainsi, il y a bien un sens en lequel il est juste de dire qu'une PERSONNE, ne peut être qu'UNE personne. Subjectivement, nous sommes une seule personne, car il n'y a pas d'*effet que cela fait* d'être plusieurs personnes. Imaginons que nous soyons – objectivement – plusieurs personnes à la fois, il n'y aurait aucun moyen – subjectivement – de s'en rendre compte. Chaque personne que nous serions aurait l'intime conviction d'être la seule à être *elle-même*. Pour la bonne raison que toutes ces personnes ne sont pas reliées entre elles et qu'elles sont donc le lieu de flux de conscience disjoints. Par conséquent, tout se passerait comme si nous n'étions qu'une personne, *celle-ci*, à l'exclusion de toutes les autres. Cela n'interdit pas que nous puissions être, objectivement, plusieurs personnes, ou plutôt que la

personne que nous sommes puisse être plusieurs individus. De même que des processus physiques représentant une expérience du vert n'ont pas besoin d'être eux-mêmes verts.

Il semble qu'un critère d'économie ontologique devrait nous conduire à considérer que, puisque nous avons subjectivement l'impression de n'être qu'une seule personne, alors c'est effectivement le cas, objectivement. Mais ce qui semble subjectivement vrai n'est pas nécessairement objectivement vrai. Ainsi, subjectivement, frapper quelqu'un d'autre que soi-même n'occasionne pas de douleur. En effet, si nous frappons quelqu'un d'autre que nous-même, nous n'expérimentons pas de douleur, nous n'expérimentons pas *sa* douleur, juste *ses* grimaces. Pourtant, qui doute sérieusement qu'autrui a mal lorsqu'on le frappe ? Si nous pouvions relier notre cerveau au sien, nous pourrions expérimenter la douleur d'autrui aussi directement que nous expérimentons les grimaces associées à son expérience douloureuse. Mais un aveugle pourrait douter aussi qu'autrui grimace quand on le frappe, puisqu'il ne verrait pas ses grimaces. Nous sommes, face à la douleur d'autrui, un peu comme des aveugles face à ses grimaces.

## *H. Intuition n° 8 : L'indexical essentiel.*

*« On peut avoir conscience de soi en première personne ou en troisième personne. Dans les deux cas, c'est bien la même réalité que l'on perçoit, mais on ne la perçoit pas de la même façon, et cela n'a pas les mêmes conséquences pratiques. Selon que l'on se pensera en première ou troisième personne, notre comportement pourra différer, de même que notre comportement peut différer selon que l'on perçoit un même état du monde extérieur comme triste ou joyeux. Ainsi, "[s]i je sais qu'un pantalon est en feu, j'aurais simplement accru ma science. Mais si je sais que j'ai le pantalon qui brûle, je vais me soucier d'autre chose que de ma science". Il semble donc qu'on ne puisse réduire le point de vue en première personne au point de*

*vue en troisième personne sans perdre en même temps une caractéristique essentielle du point de vue en première personne : son indexicalité. »*

Critique :

Cette intuition est proche de la n°6.

Que des descriptions ne puissent se réduire l'une à l'autre n'implique pas que ces descriptions ne puissent être des descriptions d'une même réalité. Il ne faut pas confondre la carte et le territoire. Chaque personne a un accès globalement plus direct à ses propres états corporels ou psychiques qu'à ceux d'autrui. C'est ce qui donne l'impression qu'il y a une dimension indexicale essentielle au sein du sujet. Or, supposons que nous ressentions une partie de notre corps comme étranger (cela arrive dans certaines pathologies), que nous la percevions comme s'il appartenait à quelqu'un d'autre, nous nous en désintéresserions sans doute. Inversement, si nous pouvions ressentir quelqu'un d'autre « de l'intérieur », exactement comme nous nous ressentons nous-mêmes habituellement, et que nous puissions agir à travers lui, exactement comme nous agissons naturellement à travers nous, cela n'aurait-il pas comme conséquence que nous nous sentirions concerné par son devenir au même titre que nous nous sentons concerné par le nôtre ? N'est-ce pas d'ailleurs ce que l'on observe dans la *sympathie* entre deux personnes ?

Ainsi, l'indexicalité ne semble pas être une propriété mystérieuse d'un sujet transcendant l'univers objectif, mais une propriété naturelle de la conscience, liée aux manières plus ou moins directes et diverses que nous avons d'expérimenter les émotions, pensées, volitions, etc. selon qu'elles sont les nôtres ou celles d'autrui.

## I. Intuition n°9 : L'irréductible libre-arbitre

*« Le libre-arbitre implique que le sujet est irréductible à un quelconque processus physique causalement déterminé et est donc une réalité à part, transcendant l'ordre naturel des choses. En effet, ce qui est causalement déterminé ne peut pas être libre. On n'aurait pas pu avoir fait autre chose que ce que l'on a effectivement fait car ce que l'on a effectivement fait est nécessité par les états antérieurs du monde. Mais de plus, la cohérence de la notion même de libre-arbitre implique que le sujet transcende l'ordre naturel des choses en s'identifiant au libre-arbitre. Comment pourrait-on dire que le sujet se détermine lui-même, autrement ? Si le sujet avait le libre-arbitre et n'était pas le libre-arbitre, de même que nous avons des vêtements, sans être des vêtements, alors le sujet ne serait paradoxalement pas libre, car il serait déterminé par cette « chose » qu'il aurait : le libre-arbitre. Pour que nous soyons véritablement libre, nous ne devons pas simplement avoir un libre-arbitre, nous devons être ce même libre-arbitre. Ainsi, chaque action libre que nous effectuons est une manifestation directe de nous-même en tant que sujet. »*

Critique :

Une première critique, simple, radicale mais difficile à admettre de cette intuition selon laquelle le libre-arbitre prouverait l'irréductibilité du sujet à un quelconque processus physique causalement déterminé consiste à mettre en doute le fait même que nous aurions ou serions un libre-arbitre. Nous aurions seulement l'impression (souvent tenace) d'en avoir un. Ainsi, pour Spinoza, la liberté n'existe pas et « L'illusion de la liberté vient de la conscience de notre action et de l'ignorance des causes qui nous font agir. »[82]. La philosophie de Spinoza défend un déterminisme absolu de l'Univers. Laplace[83] a formulé

[82] SPINOZA, Baruch, *Ethique II*, proposition XXXV, scolie

[83] LAPLACE, Pierre-Simon. *Essai Philosophique sur les Probabilités.*

scientifiquement ce déterminisme et l'a illustré par une expérience de pensée célèbre (qui est restée à la postérité sous le nom d' « expérience de pensée du démon de Laplace ») :

> Nous devons envisager l'état présent de l'univers comme l'effet de son état antérieur, et comme la cause de celui qui va suivre. Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux. L'esprit humain offre, dans la perfection qu'il a su donner à l'astronomie, une faible esquisse de cette intelligence. Ses découvertes en mécanique et en géométrie, jointes à celles de la pesanteur universelle, l'ont mis à portée de comprendre dans les mêmes expressions analytiques les états passés et futurs du système du monde. En appliquant la même méthode à quelques autres objets de ses connaissances, il est parvenu à ramener à des lois générales les phénomènes observés, et à prévoir ceux que les circonstances données doivent faire éclore.

Autrement dit, les choses n'auraient pas pu se passer autrement qu'elles se sont effectivement passées. Le libre arbitre est un leurre.

Le déterminisme pourrait cependant n'être que régional, et ne pas forcément concerner tous les systèmes. Une part de vrai hasard métaphysique pourrait intervenir. A *priori*, ce hasard pourrait être de deux types. Le hasard des statistiques et des probabilités n'est pas prévisible à l'échelle de l'événement, mais l'est à l'échelle de la collection d'événements. Il suit en général une loi dite Normale. Il s'agit donc d'un hasard qui n'est pas absolu, d'un hasard que l'on pourrait qualifier de modéré. On pourrait cependant concevoir un hasard beaucoup plus sauvage, qui ne serait modélisable par aucune loi statistique.

En physique des particules, les inégalités d'Heisenberg semblent montrer qu'une part de hasard régit le monde des particules dans la mesure où il est impossible de déterminer avec précision à la fois la position et la vitesse d'un électron. Toute précision concernant un des deux paramètres s'accompagne d'une imprécision sur le second.

Notons toutefois que cela ne prouve pas que tout soit indéterminé. L'absence de prévisibilité ne signifie pas l'absence de déterminisme. Elle pourrait venir davantage des limites de notre esprit que d'une indétermination réelle du futur. En effet, un hasard modéré (tel que défini plus haut) ne serait pas différenciable du hasard épistémologique qui est compatible avec un déterminisme universel, lié à notre ignorance et dont l'existence n'est pas remise en cause. Un hasard sauvage qui n'obéirait à aucune loi statistique serait en revanche beaucoup plus difficile à comprendre dans une perspective purement déterministe. Mais un tel type de hasard a-t-il déjà été mis en évidence ? Il est difficile de se prononcer. Le problème avec ce type de hasard c'est qu'étant donné sa nature absolument imprévisible, ses éventuelles manifestations seraient non-reproductibles. Or l'épistémologie actuelle pose le critère de reproductibilité en réquisit indispensable à l'admission de l'existence d'un phénomène. Reste qu'il existe effectivement un épais dossier riche en anecdotes plus ou moins bizarres relatant des phénomènes spontanés défiant *a priori* toute tentative de systématisation (miracles, apparitions spirituelles, etc.). Mais l'hypothèse de la fraude, de l'illusion ou de l'hallucination ne peut être écarté dans la plupart des cas répertoriés. De plus, à supposer que ces phénomènes et leur caractère inexplicable soient admis, l'hypothèse d'un hasard sauvage serait-elle la seule explication par défaut envisageable ? C'est très discutable. Ce qui paraît inexplicable pourrait en fait avoir une explication qui ne soit pas accessible à notre investigation, par exemple une explication située dans un autre univers avec une dimension supplémentaire dont notre univers ne

serait qu'une partie. Des objets 4D traversant notre univers 3D se manifesteraient par l'apparition et la disparition progressive d'objets 3D, sans qu'une cause située dans notre univers 3D puisse être identifiée et sans que ce phénomène puisse être reproduit en opérant à travers notre univers 3D. Phénoménologiquement, on aurait l'impression d'un phénomène hasardeux au sens le plus fort, alors qu'en fait, il s'agirait d'un phénomène tout à fait déterministe mais dont la cause ne nous serait pas accessible[84].

Admettons toutefois, pour les besoins de la discussion, qu'un hasard métaphysique existe vraiment. Supposons en outre qu'il s'en trouve dans la volition et l'action humaine. Cette part de hasard peut-elle être identifiée à ce que l'on appelle le libre-arbitre ? Il semble que non. Si on veut ou agit par hasard, ce n'est pas librement. Intuitivement, si on poignarde son meilleur ami par hasard, on a plutôt envie d'invoquer un accès de folie qu'un acte libre… De même, une personne qui prendrait toutes ses décisions en tirant aux dés parmi les possibilités s'offrant à elle n'agirait pas librement.

Qu'il soit difficile de se former un concept, qui plus est un concept cohérent du libre-arbitre, ne va pas sans jeter le doute sur l'existence de ce libre-arbitre. Si la liberté, en soi, ne pose pas de problème particulier dès lors qu'on la définit comme le fait de pouvoir agir sans contrainte, c'est-à-dire, d'avoir la capacité de faire ce que l'on veut, le problème se corse à partir du moment où l'on examine le libre-arbitre, encore appelé « volonté libre ». En effet, que doit-on entendre par là ? Le fait de pouvoir vouloir sans contrainte ? De vouloir ce que l'on veut ? Mais n'est-ce pas une absurdité d'imaginer que l'on pourrait vouloir autre chose que ce que l'on veut ? Et par conséquent, n'est-ce pas une tautologie

84 Afin de mieux comprendre ce point, on pourra lire avec intérêt ABBOTT, Edwin Abbott. *Flatland : A Romance of Many Dimensions*.

que de dire que l'on peut vouloir ce que l'on veut ? A supposer qu'il existe une volonté de deuxième ordre qui détermine ce que l'on appelle « notre volonté », alors cette dernière ne serait plus libre, et la question resterait entière : cette volonté de deuxième ordre est-elle libre ?

La difficulté à définir le libre-arbitre ne doit pas nous décourager. Il est peut-être inhérent à sa nature de ne pouvoir se laisser définir facilement. Le libre-arbitre est imprévisible et peut prendre des tas de formes. Comment ce qui est imprévisible et polymorphe pourrait-il se laisser enfermer dans une définition ? Ne dit-on pas de quelqu'un qu'il est original, singulier et qu'il a de la personnalité lorsque justement, on ne peut pas le définir, le cataloguer dans un type bien précis ?

Pour la même raison, le libre-arbitre ne peut être scientifiquement prouvé. Ce qui est scientifiquement prouvable est objectivement descriptible et est régulier dans sa manifestation, une fois les conditions initiales posées. C'est donc tout le contraire du libre-arbitre ! Un tel libre-arbitre existe-t-il vraiment ? Le problème est le même que pour le hasard métaphysique. Si un libre-arbitre existe, il semble indiscernable du « bruit de fond » que serait une volonté déterminée et ignorante de sa nature déterminée. Et même si le comportement d'un individu en particulier est difficile à prévoir, celui d'une masse d'individus l'est beaucoup moins et semble obéir à des lois statistiques. En témoigne le fait que le statut scientifique de la sociologie est moins sujet à caution que celui de la psychologie, du moins les branches de la psychologie qui sont centrées sur l'individu. Auguste Comte considérait la sociologie comme une science (une « physique sociale »). A contrario, la psychologie était absente de sa fameuse classification des sciences.

Pour qu'un libre-arbitre puisse sortir du « bruit de fond », il faudrait déjà qu'il soit proprement étonnant. Ce serait une condition nécessaire mais pas suffisante à la reconnaissance de sa nature. En effet, le problème est le même que pour le hasard « sauvage » et métaphysique que nous avons défini plus haut. Des récits d'actions ou de volontés héroïques ou *a contrario* particulièrement odieuses existent. Mais, à supposer qu'on puisse les authentifier et même en préciser le caractère inexplicable, l'existence d'un libre-arbitre fort ne serait pas la seule interprétation possible étant donné qu'on ne sait pas tout ce qui pourrait éventuellement nous déterminer de manière « occulte ».

A défaut de pouvoir correctement définir le libre-arbitre, nous pouvons peut-être le montrer, en esquisser la notion à partir de quelques situations que l'on qualifie de libres. Dans quels cas dirait-on avec assurance que le libre-arbitre était en jeu ?

Le libre-arbitre est la capacité supposée qu'a le sujet de se déterminer par lui-même, échappant ainsi au déterminisme physique. En effet, quand le libre-arbitre agit, l'action et/ou la volonté du sujet à un temps t2 ne dépend plus entièrement de l'état du monde en t1. Le libre-arbitre doit être distingué du simple hasard. Quand on agit au hasard, on n'agit pas librement, nous l'avons vu. Si une personne lançait une pièce en l'air avant de prendre toutes ses décisions, on ne dirait pas d'elle qu'elle est libre. Quand on agirait librement, on continuerait de suivre une certaine norme (tout au moins on essayerait) et notre action serait donc en partie justifiable, quelque soit la décision que l'on prendrait. Le libre-arbitre est donc capacité à faire des choix délibérés et il est générateur de sens. Le choix de la norme n'est pas forcément libre, lui, cela dit, car *in fine*, on voit mal comment le choix de la norme de toutes les normes pourrait suivre une norme…

Ensuite, le libre-arbitre est créateur, ou plutôt : toute création proprement dite ne peut être que libre. On voit mal comment on pourrait véritablement créer si on ne fait qu'appliquer une méthode, ou comment le processus consistant à faire travailler une machine qui génèrerait aléatoirement toutes les œuvres d'arts possibles jusqu'à ce qu'on en trouve une à notre goût serait créatif… La liberté requiert la spontanéité, et ce qu'on appelle l'*insight*, qui est la perception directe d'une solution a un problème ouvert (i.e. dont le nombre de solutions possibles n'est pas défini), sans passer par une démarche analytique.

Enfin, l'action libre s'accompagnerait d'un sentiment que, toutes choses égales par ailleurs, on aurait pu (et parfois aussi : on aurait dû) faire autre chose.

En termes de mondes possibles, on peut définir une action ayant lieu en t2 comme étant libre si (1) il existe au moins un monde possible où elle n'a pas lieu, quand bien même l'état du monde en t1 est le même que dans le monde actuel ; (2) dans le monde actuel et dans tous les mondes possibles où l'état du monde en t1 est le même que dans le monde actuel, l'action réalisée est justifiable à l'égard d'une certaine norme et s'accompagne du sentiment que toutes choses égales par ailleurs, on aurait pu (et parfois aussi : on aurait dû) faire/vouloir autre chose ; (3) l'action actuelle et ses contreparties sont spontanées, elles ne sont pas le fruit d'une démarche analytique, de l'application d'une recette...

On pourrait dire aussi – négativement – qu'une action libre, est une action qui n'est ni fortuite ni causalement déterminée.

Il y a de bonnes raisons ou plutôt de bons motifs de croire au libre-arbitre, même s'il est difficile à définir et impossible à prouver scientifiquement. Kant en a donné quelques-

unes. Le libre-arbitre est un postulat de la raison pratique[85] : c'est une condition transcendantale de la morale. La morale, la responsabilité, le mérite, la culpabilité, etc. ont-ils un sens pour des êtres entièrement déterminés causalement ? Si nous ne sommes pas libres, alors aucun assassin n'est responsable de ses meurtres : il ne pouvait pas faire autrement que de les commettre ! Et pourquoi alors éprouver de la haine pour lui ? On pourrait objecter que nous ne sommes pas libres de ne pas en éprouver. Nous répondrions qu'une croyance en l'absence de libre-arbitre détermine causalement que l'on ne peut pas éprouver de la haine pour un assassin, mais plutôt de la compassion. De la même façon que l'on n'éprouvera pas de haine mais de la compassion pour un handicapé, un arriéré mental, un malade, etc. Pourtant, combien de personnes convaincues par l'inexistence du libre-arbitre en tirent-elles toutes les conséquences pratiques ? Bien peu semble-t-il, et heureusement !

La volonté libre serait-elle indéfinissable, à part de manière circulaire et tautologique, cela ne signifierait pas pour autant qu'elle n'existe pas. En effet, nous avons une forte intuition de cette existence même si nous nous trouvons désemparé lorsque vient le moment de la circonscrire. « Qu'est-ce donc que le temps ? Si personne ne me le demande, je le sais; mais si on me le demande et que je veuille l'expliquer, je ne le sais plus. », écrivait Saint Augustin[86]. Pourtant, il semble difficile de nier l'existence du temps qui passe[87]. La

---

85 KANT, Emmanuel, *Critique de la Raison Pratique*.

86 AUGUSTIN, Saint. *Confessions*, XI, 14

87 D'autres exemples de choses indéfinissables et dont l'existence est ou semble en tout cas difficilement contestable (dans certains cas, on peut montrer des exemples de ces choses, et donc les définir de manière ostensive) : la définition, le sens ou la signification (si on pouvait donner une définition correcte de définition, de sens ou de signification, cette définition ou son contenu serait inclus dans son extension, autrement dit, cette définition ou son contenu serait un exemple de lui-même, ce qui est paradoxal), l'existence ou l'être, le langage, les mots, les lettres, etc.

même chose pourrait être dite pour le libre arbitre. Comme l'écoulement du temps, il est, nous l'avons vu, un présupposé implicite de toutes nos actions. Si quelqu'un se comportait comme si le libre-arbitre n'existait pas, alors il pourrait tout aussi bien s'allonger en plein milieu d'une route et attendre qu'une voiture l'écrase ; rester assis sans rien faire, se laisser mourir de faim, tuer la première personne qu'il rencontre, s'adonner sans retenue à l'alcool, au tabac et à la drogue. Ce ne serait pas sans être inquiété, car il va de soi que la société ne le laisserait pas faire tout cela et que lui-même pourrait être malheureux des conséquences de ses actes. Mais ce serait sans se culpabiliser de les avoir accomplis puisqu'il n'y avait pas d'autres possibilités de déroulement des événements que ce qui s'est effectivement produit. Il pourrait se dire, tout en reconnaissant que d'autres choix auraient pu être plus judicieux, qu'il n'y a pas lieu qu'il se reproche quoi que ce soit, car il n'aurait pas pu choisir autre chose que ce qu'il a effectivement choisi, il n'aurait pas pu vouloir autre chose que ce qu'il a effectivement voulu.

A ceux qui pourraient objecter que sa non croyance au libre arbitre aurait pu aller de paire avec un sentiment de culpabilité, ce sentiment étant lui aussi déterminé, il suffirait de faire remarquer que ce sentiment serait incompatible avec une non croyance au libre arbitre pleinement assumée. Un sentiment de culpabilité arrivant malgré une croyance profonde en l'inexistence du libre-arbitre, nécessiterait de pouvoir s'arracher du déterminisme naturel conduisant à ne pas se sentir coupable de ses actes lorsque l'on ne croit pas les avoir librement voulu, et donc, supposerait quelque part de souscrire à l'idée d'un certain libre arbitre.

De même, une société où tout le monde croirait fortement en l'inexistence du libre arbitre pourrait très bien avoir un système pénal et judiciaire, réprimer le crime, exiger la

justice, punir les malfrats, etc. sans se contredire. Ce serait dans un pur but de sécurité et d'amélioration du comportement des auteurs d'actes répréhensibles. Mais cela ne pourrait pas être dans un but revanchard ou pour soulager émotionnellement les victimes. On prescrirait plutôt à ces dernières des séances de psychothérapie cognitive où on leur montrerait l'incompatibilité du sentiment de vengeance avec le fait admis que le libre-arbitre n'existe pas. Car en effet, commettre le mal ne serait plus alors qu'une question de maladie à soigner, voire un pur incident. Comment rationnellement en vouloir à quelqu'un d'être malade ?

Le sentiment de liberté est quelque chose de très profondément enfoui en chacun, et même ceux qui pensent que le devenir est entièrement déterminé vont très rarement jusqu'à assumer cette position en pratique. La plupart des gens font comme si le libre-arbitre existait, et cela même s'ils prétendent croire fortement qu'il n'existe pas.

Une autre raison ou un autre motif de croire à l'existence du libre-arbitre est que nous avons été très probablement sélectionnés par l'évolution pour y croire, tellement ce sentiment est naturel, universel et profondément enfoui en nous. Si tel est le cas, c'est un motif supplémentaire pour penser que la croyance au libre-arbitre est nécessaire à notre survie et à notre pérennité. Se refuser à avoir cette croyance revient à se refuser à accomplir une certaine fonction biologique quelconque : digestion, respiration, perception, cognition, locomotion, etc. Or, il va de soi que s'amputer d'une fonction biologique nous rend moins adapté à notre environnement, s'oppose à notre aspiration naturelle à la vie et à la reproduction.

On peut se demander toutefois si ces motifs éthiques et biologiques de croire au libre-arbitre sont épistémiquement valables. N'est-ce pas de la plus pure *wishful thinking* ?

La question est délicate… En prenant ses désirs pour des réalités, la plupart du temps, on se trompe, même si on peut occasionnellement tomber juste, et on comprend pourquoi : on ne voit pas pourquoi la réalité, qui est (pour autant que nous en sachions) objective, devrait s'infléchir sous l'effet de notre désir qui, lui, est le plus souvent subjectif. Le plus souvent. Mais que penser d'un désir puissant et partagé par tous ? Un désir si puissant de croire, que toute démonstration rationnelle échoue à mettre en doute cette croyance.

Si un homme croit que toutes les femmes sont forcément conquises par son charme, la réalité aura bien vite fait de corriger sa croyance. Et si ce n'est pas le cas, il passera pour un incorrigible érotomane aux yeux de la plupart des gens… Si un homme croit que le libre arbitre existe, non seulement la réalité ne réfutera pas sa croyance, mais elle le rendra heureux (n'est-il pas déprimant de penser que nous serions des automates, si ce n'est que ça nous déculpabilise de nos actions ? Mais pour quelqu'un de droit, n'est-ce pas déprimant de se dire qu'il n'a aucun mérite réel ?) et, de plus, il trouvera plein de gens autour de lui qui partageront sa croyance et l'affermiront dans celle-ci. Observons à présent une autre croyance bien implantée, à vrai dire tant et si bien implantée que l'on a peine à la considérer comme une croyance : la croyance en la raison.

Si une personne croit que l'univers est rationnel, on ne peut pas dire non plus, comme pour la croyance au libre-arbitre, que la réalité viendra le contredire. Cette croyance le rendra heureux. Parfois c'est contestable (il ne faut pas abuser des bonnes choses !), et on

peut sombrer dans le rationalisme morbide, trouvant un ennui profond à un univers où tout obéi à une logique implacable… Cela dit, généralement, il est très plaisant de faire fonctionner son esprit de manière logique (les passionnés de problèmes logiques comprendront !). Et la logique, alliée à l'expérience, a permis à l'humanité, en une centaine d'années, un progrès indéniable. Beaucoup plus important que celui accumulé depuis plusieurs millénaires que dure l'histoire humaine. Progrès qui se manifeste notamment par une augmentation de notre qualité de vie. Enfin, l'adepte de la raison trouvera autour de lui près de 6 milliards d'individus partageant la même croyance et l'affermissant dans ses convictions !

Mais n'avons-nous pas d'autres raisons de croire en la raison ? Des raisons valables ? Comment pourrait-il jamais y en avoir ? Si une raison valable de croire en une chose implique que cette croyance soit justifiée par la raison, alors nous n'avons aucune raison valable de croire en la raison. Comment la raison pourrait-elle s'auto-justifier ? Et si l'on admet que tous les cercles vicieux ne sont pas des absurdités (pourquoi pas), alors n'est-ce pas là la porte ouverte à l'auto-justification de tous les systèmes philosophiques que l'on voudra ?

Il est vrai que si quelqu'un nie ou doute que les lois régissant l'univers observable sont rationnelles, il passera pour un excentrique. Toutefois, rien dans la réalité ne viendrait forcément le contredire. Au lieu de croire que les pommes tombent des arbres sous l'effet d'une force (la force de gravitation) $F=G.m_1.m_2/d^2$ où G est la constante de gravitation, $m_1$ la masse de la pomme, $m_2$ la masse de la Terre et d la distance séparant le centre de gravité de la pomme du centre de gravité de la Terre. On pourrait aussi croire que la chute des pommes est en fait l'œuvre d'un malin génie invisible qui s'arrange pour que toutes nos

mesures empiriques de la chute des pommes soient compatibles avec cette formule, et se passer totalement de la notion de force. Plus concrètement, certains pans de la physique comme la mécanique analytique[88] se passent dans les faits de la notion de force.

De même, concernant le libre arbitre, c'est bien plutôt ceux qui le nient ou le mettent en doute que ceux qui l'affirment qui passent pour des excentriques. Bien sûr, quelqu'un qui pense que le libre-arbitre n'existe pas ne sera pas plus contredit par la réalité que celui qui pense qu'il existe. De même que précédemment, celui qui pense que les lois de l'univers sont irrationnelles ne sera pas plus contredit par la réalité que celui qui pense qu'elles sont rationnelles. Il semble toutefois que la croyance au libre-arbitre fasse partie de ces « croyances proprement basiques » telles que les définit le philosophe et apologiste chrétien Alvin Plantinga[89], et qui comprennent aussi : la croyance en sa propre existence et en son propre passé, la croyance en l'existence du monde extérieur, la croyance en l'existence des autres esprits (*belief in the existence of other minds*), etc.

Mais on pourrait aussi rajouter à ces croyances proprement basiques : la confiance que l'on accorde à notre raison et à notre expérience. Les prédispositions cognitives à l'usage de la raison et de l'expérience, ainsi que la confiance qu'on leur accorde sont sans doute, si on en croit justement ce que nous enseignent la raison et l'expérience, le fruit d'une évolution néo-darwiniste. Celle-ci sélectionne des fonctions adaptées aux contraintes de l'environnement et non des fonctions parfaites. Par conséquent, une personne empirico-

---

88 La mécanique analytique regroupe diverses formulations non-newtoniennes (se passant du concept de force) de la mécanique classique, telles la mécanique hamiltonienne et la mécanique lagrangienne.

89 PLANTINGA, Alvin. *God and Other Minds: A Study of the Rational Justification of Belief in God.*

rationaliste, qui semble par là même étrangère à toute foi, a en fait une certaine foi (au sens de confiance) en la fiabilité de la raison et de l'expérience.

Les prédispositions cognitives à l'usage confiant de la raison et de l'expérience, se retrouvent donc, d'après le néodarwinisme, dans une situation comparable aux prédispositions cognitives à l'usage confiant de ce que l'on appellera faute de mieux, l'*intuition*. Le degré de confiance effective que l'on accordera à la raison et à l'expérience sera donc aussi (peu) justifié que le degré de confiance effectif que l'on accordera à l'*intuition*[90]. Par conséquent, le degré de croyance que l'on aura envers par exemple, la thèse du déterminisme (issu, admettrons-nous pour les besoins de l'argument, de l'usage de la

[90] On pourrait objecter que la raison et l'expérience, par rapport à l'intuition, présentent la particularité de ne fournir qu'une seule réponse à chaque question que l'on se pose sur le monde, tandis que l'intuition proposerait (notamment, selon les personnes) plusieurs solutions entre lesquelles on ne saurait trancher, et que la raison et l'expérience sont bien mieux partagés entre tous les hommes, qu'elles peuvent donc prétendre à une universalité à laquelle l'intuition ne peut prétendre (tout le monde n'a pas autant d'intuition, ou pas la même). Sans doute que le degré de certitude que l'on peut accorder aux croyances intuitives ne vaut pas celui que l'on peut accorder aux croyances empirico-rationnelles. Pour autant, on ne peut lui refuser toute valeur épistémique. Si l'intuition confond parfois ceux qui en font usage, cela peut venir d'un manque d'élaboration de cet outil, plus que d'une faille essentielle de celui-ci. La raison et l'expérience ont aussi divers degrés d'élaboration. Dans le sens commun, raison et expérience sont utilisés d'une manière frustre, rudimentaire et partiellement inconsciente, ce qui aboutit à des erreurs, des paradoxes et des paralogismes. Mais cela ne signifie pas que l'expérience et la raison sont essentiellement trompeuses. A travers la méthode scientifique, qui en constitue une formalisation et une sophistication, elles sont largement moins sujettes à ces égarements (toujours possibles cependant). On peut donc, peut-être concevoir une « méthode intuitive » ou une « intuition normalisée » qui pourrait venir palier aux problèmes de l'intuition brute de tout un chacun. Enfin, notons que la raison et l'expérience ne sont pas si partagées que l'on pourrait le penser, puisque des études en psychologie cognitive montrent tous les biais de raisonnement qui affectent la plupart des gens, avec parfois une force insoupçonnée, et même dans des situations où une certaine concentration est sollicitée. De ce que la logique soit normalisée et codifiée, on ne doit en conclure que c'est le cas de notre logique naturelle, et l'éducation joue pour une bonne part dans le développement chez l'enfant et l'adulte d'une pensée logique objective. Cette même éducation, malheureusement, n'apprend nulle part à correctement utiliser son intuition en vue de développer ce que l'on pourrait appeler une « intuition objective ».

raison et de l'expérience) sera aussi (peu) justifié que le degré de croyance que l'on aura envers par exemple, la thèse du libre-arbitre.

Il faut reconnaître, contre certains apologistes chrétiens comme Alvin Plantinga[91], que le degré de confiance accordé à la démarche scientifique par ses adeptes, est rarement de la même ampleur que la Foi absolue de certains chrétiens envers les « vérités révélées ». On peut accorder à la science une valeur épistémique relative (la science nous permet de connaître les phénomènes, et non pas ce qui se cache éventuellement derrière) voire même seulement une valeur pragmatique (la science nous donne accès à des « vérités » utiles à notre survie, notre bonheur, etc.). Selon les cas, on sera donc justifié à accorder une valeur épistémique absolue, relative, ou seulement une valeur pragmatique à l'*intuition*. Par conséquent, les tentatives de réfuter le naturalisme au profit du théisme ne peuvent au mieux que réfuter toute métaphysique au profit d'un sage agnosticisme. Toutefois cet agnosticisme autorise un fidéisme de type wittgensteinien, où le langage tant religieux que scientifique doit se comprendre et s'évaluer en référence à un certain cadre, un certain jeu de langage (qui caractérise une certaine forme de vie), dont le choix ne peut pas lui-même être justifié de l'intérieur : on ne peut pas justifier scientifiquement la validité de la démarche scientifique, pas plus qu'on ne peut justifier religieusement la validité de la démarche religieuse. Nous avons pris l'exemple de la pensée religieuse, car c'est surtout par rapport à elle que l'on parle de fidéisme (et donc aussi de fidéisme wittgensteinien), mais toute autre démarche intellectuelle incluant, comme axiome, présupposé, théorème ou résultat, le libre-arbitre, ne peut se justifier de l'intérieur, sans que cela implique pour autant qu'elle doive être considérée comme métaphysiquement fausse.

---

91 PLANTINGA, Alvin, "An evolutionary argument against naturalism".

En résumé, si l'on accorde une valeur épistémique élevée à la démarche scientifique, on est paradoxalement justifié à croire au libre-arbitre, si l'on accorde une valeur épistémique basse voire une simple valeur pragmatique à la démarche scientifique, alors on n'est pas justifié à rejeter la thèse du libre-arbitre.

La charge de la preuve reviendrait donc plutôt à celui qui nie plutôt qu'à celui qui affirme le libre-arbitre. Notons aussi que parmi les gens qui prétendent nier le libre-arbitre, bien peu semblent convaincus par leurs propres arguments, puisque leur comportement, au jour le jour, ne diffère pas sensiblement de celui des gens qui prétendent croire au libre-arbitre. Or, s'il est déjà étonnant qu'une croyance n'ait aucune répercussion pratique, et paraisse n'être qu'une pure fantaisie intellectuelle, il est encore plus étonnant qu'une croyance au sujet du libre-arbitre, donc au sujet de la nature même de nos comportements, n'ait aucune répercussion sur nos comportements ! Notre homme qui croit que toutes les femmes sont forcément conquises par son charme serait-il cohérent avec lui-même si, en même temps, il était intimidé à leur parler, souffrait d'angoisse de leur déplaire, n'avait aucune confiance en lui pour les aborder ? Serions-nous tenté de dire de lui qu'il croit *réellement* que toutes les femmes sont forcément conquises par son charme ? A la limite, nous pourrions lui concéder qu'il croit le croire, qu'il le croit consciemment mais pas au plus profond de lui-même. Et c'est bien là le même problème avec les personnes qui rejettent l'idée du libre-arbitre : leur comportement n'est pas en adéquation avec leur prétendu rejet. Elles continuent de ressentir de la culpabilité, d'éprouver de la haine, d'avoir du remord et des regrets, etc. Quelqu'un qui croirait qu'un malin génie est responsable de la chute des pommes devrait normalement être anxieux en permanence car un malin génie est imprévisible, très puissant et, de plus, pourrait être fort mécontent que quelqu'un soit au courant de son existence. Il pourrait donc réserver un bien mauvais sort à toute personne

qu'il jugerait indésirable. Ces conséquences bien pratiques de la croyance en un malin génie nous permettent de conclure que toute personne qui prétendrait donner crédit à cette hypothèse tout en menant une vie tranquille et rassurée serait un menteur ou un inconscient !

Nous terminerons cet exposé des motifs à croire en un libre-arbitre par un argument proposé par le philosophe Michael Huemer, dans son article *A proof of Free Will*[92]. Une preuve du libre-arbitre, la démonstration de Huemer n'en est sans doute pas une, malgré ce que pourrait suggérer le titre de l'article. Mais c'est en tout cas un argument assez intéressant en faveur de son existence. Son raisonnement est le suivant :

1. Par rapport au problème du libre arbitre, nous devons ne pas entretenir de fausses croyances. (Prémisse)[93]

2. Quoi que ce soit qui doive être fait peut être fait.[94] (Prémisse)

3. Si le déterminisme est vrai, alors quoi que ce soit qui peut être fait, est fait.[95] (Prémisse)

4. Je crois au libre-arbitre[96]. (Prémisse)

---

92 HUEMER, Michael, "A proof of Free Will".

93 C'est bien entendu discutable, mais de prime abord de bon sens.

94 En effet, on voit mal par exemple comment une personne amputée des deux pieds pourrait essuyer ses pieds avant d'entrer dans une maison, même s'il est écrit sur la porte : « Prière d'essuyer vos pieds avant d'entrer ».

95 Cette proposition est logiquement équivalente à : « Si le déterminisme est vrai, rien de ce qui n'est pas fait n'aurait pu être fait. »

5. Par rapport au problème du libre-arbitre, nous pouvons ne pas entretenir de fausses croyances. (De 1 et 2)

6. Si le déterminisme est vrai, alors par rapport au problème du libre-arbitre, nous n'entretenons pas de fausses croyances. (De 3 et 5)

7. Si le déterminisme est vrai, alors le libre-arbitre existe. (De 6 et 4)

8. Le libre-arbitre existe. (De 7)[97]

Si se passer de l'hypothèse du libre-arbitre n'est pas une bonne idée, alors que répondre au dualiste qui y voit une preuve de l'irréductibilité du sujet à un système causalement déterminé ?

Sur le problème des rapports entre liberté et déterminisme, on distingue deux approches. Pour l'approche incompatibiliste, la liberté et le déterminisme entrent en conflit : il s'ensuit que si le déterminisme est absolu, la liberté ne peut exister. Par conséquent, ou bien le déterminisme est absolument vrai, et la liberté n'existe pas, ou bien il est, au mieux, relatif, et la liberté peut exister. L'approche incompatibiliste semble pleine de bon sens, à la limite de la lapalissade. Pourtant une autre approche existe, l'approche

---

[96] L'auteur, dans son article, dit plutôt : « Je crois en la MFT » où MFT signifie : *Minimal Free-Will Thesis* (Thèse du Libre-Arbitre Minimal) qu'il définie comme suit : *« [A]u moins de temps en temps, quelqu'un a plus d'une ligne de conduite qu'il peut effectuer. […] C'est le minimum qui doit être vrai pour que l'on puisse dire qu'il y a du libre-arbitre. »* (il reprend cette terminologie de l'article de Peter Van Inwagen : *An Essay on Free Will* [Oxford: Clarendon Press, 1983], p. 91). La MFT correspondrait plutôt à ce que nous appelons « indéterminisme ».

[97] En toute rigueur, il manque une prémisse qui serait : « Le déterminisme est vrai ».

compatibiliste, qui soutient au contraire que la liberté peut exister dans un monde parfaitement déterministe, voire que le déterminisme est une condition nécessaire à l'exercice de la liberté. Il apparaît toutefois qu'une telle position implique une redéfinition de la liberté gommant sa dimension métaphysique et insistant sur sa dimension épistémologique. La liberté existe car nous ne pouvons pas, en principe, connaître toutes les causes qui nous déterminent, et les connaître dans les détails des détails des détails. Les compatibilistes s'intéressent donc à une liberté émergeant de phénomènes chaotiques, parfaitement déterministes mais parfaitement indéterminables à cause de nos limitations cognitives. On peut défendre une autre forme de compatibilisme : le monde, pour nous, est constitué de montagnes, de vallées, de rivières, d'animaux, de plantes, de chaises, de tables, de rouge, de bleu, de froid, de chaud, de dur, de mou, de sociétés, d'états, de peuples, d'organisations, etc. Il est également constitué de libre-arbitre. Toutefois, le mode d'existence de toutes ces entités est relatif. Dans l'absolu, il n'y a que des particules physiques élémentaires en interaction selon un mode déterministe. Ce sont elles qui constituent la vraie substance du monde. De même que l'on a le droit de dire que les montagnes ou le rouge existent, on a donc le droit de dire que le libre-arbitre existe, et en disant cela, on ne raconte pas de bêtise, mais il faut garder à l'esprit que ces existants sont relatifs.

Pour notre part, nous suggérerons une position intermédiaire entre l'incompatibilisme et le compatibilisme. Quand on pense au libre-arbitre, on pense à une capacité du sujet dont le degré de réalité serait au moins égal à celui du déterminisme causal de la matière. Si le second a une existence relative, alors le premier peut avoir une existence relative ou absolue ; mais si le déterminisme est absolument vrai, alors la liberté doit avoir une réalité absolue également.

Déterminisme et liberté ne peuvent avoir une existence absolue simultanément qu'à la condition d'imaginer un certain dualisme. Il existe une certaine portion de la réalité qui est régi par des lois strictement déterministes (appelons-la *matière*), tandis qu'une certaine autre portion de la réalité est libre (appelons-la *esprit*). L'esprit utilise la matière comme un instrument pour ses fins, de la même manière que les êtres humains utilisent des machines (automobiles, robots, ordinateurs, etc.) pour leurs fins. Pour se faire, l'esprit génère une action causale sur certains éléments matériels pour les modifier, puis cette modification se répercute sur l'ensemble du système de manière strictement déterministe. On peut se demander toutefois si cette interaction entre l'esprit et la matière ne porte pas atteinte à leur absoluité. L'esprit se relativise dans l'effort qu'il doit fournir pour maîtriser la matière. La matière se relativise dans l'indétermination qu'elle acquiert sous l'action de l'esprit.

Mais il est possible aussi que seule la liberté soit absolue, tandis que le déterminisme serait relatif. Cela est possible si par exemple, le déterminisme n'est qu'un cas particulier de la liberté. Or, si l'on voit mal comment la liberté pourrait émerger du déterminisme, en revanche, on comprend bien comment le déterminisme pourrait émerger de la liberté. Quelqu'un peut librement choisir de construire un système obéissant à des lois déterministes (une montre par exemple). Pour les philosophes déistes du siècle des Lumières, Dieu a choisi librement de créer un monde régi par des lois déterministes. Ces lois déterministes deviennent alors une expression de sa liberté. On peut aussi librement choisir de se plier à un certain nombre de règles déterministes. Si on ne pouvait pas le faire, on ne serait pas entièrement libre. Si on peut le faire, et tant que l'on s'y plie, notre conduite est déterminée et pourtant elle est libre[98].

[98] Nous laisserons de côté la question de savoir si on peut librement choisir de ne pas être libre…

Enfin, il est possible que le déterminisme et la liberté soient tous deux relatifs. Dire ceci implique : ou bien qu'il n'y a pas d'absolu, ou bien que s'il y en a un, il n'est ni libre ni déterminé. Il pourrait par exemple être indéterminé (sans toutefois être véritablement libre), ou bien partiellement indéterminé ou libre et partiellement déterminé. Dans ce dernier cas, le réel serait par exemple statistiquement déterminé (on pourrait prédire avec un bon intervalle de confiance, le comportement d'un grand groupe d'individus) mais non pas strictement déterminé (on ne pourrait pas prédire, avec un bon intervalle de confiance, le comportement d'individus isolés). Des phénomènes physiques comme la désintégration radioactive offre une très bonne image de ce que pourrait être le réel intrinsèquement. Mais on peut citer aussi les lois de masse de la sociologie. Si la sociologie, contrairement à la psychologie, fut considérée par Auguste Comte comme une science, c'est qu'elle permet d'effectuer des prédictions. La psychologie, du moins dans sa forme dite « clinique », se focalise sur un individu isolé. La liberté de cet individu rend son comportement peu prévisible. Dans un groupe humain, toutefois, les libertés individuelles se neutralisent plus ou moins, si bien que le comportement collectif d'un groupe humain s'avère bien plus prévisible dans l'ensemble que le comportement d'un seul des individus qui le constitue [99].

Dans un tel univers, dire qu'un comportement est libre ou déterminé, relèverait en bonne partie aussi du point de vue que l'on adopte pour l'étudier. Il va sans dire qu'une méthode d'investigation comme la démarche scientifique qui présuppose le déterminisme de son objet, le percevra comme déterminé par des lois causales, ou plutôt ne pourra connaître de lui que ce qui peut se décrire comme déterminé par des lois causales. En revanche, une méthode d'étude présupposant le libre-arbitre de l'être humain (humanisme,

[99] Au passage, les phénomènes de masse et la désintégration radioactive sont deux autres exemples où un comportement déterministe (quoique statistiquement) peut émerger d'un comportement (en partie) libre.

romantisme, théologie, etc.), verra dans celui-ci un libre-arbitre agissant. Encore une fois, ces descriptions incompatibles ne doivent pas nous faire croire que les réalités correspondantes sont incompatibles. Encore une fois, il faut se garder de confondre la carte et le territoire. La méthode scientifique nous dresse une certaine carte de la réalité, mais d'autres cartes existent ou pourraient exister : plus ou moins précises, plus ou moins exactes, mettant plus ou moins en valeur certains traits de la réalité au détriment d'autres, décrivant la réalité selon un certain angle, un certain point de vue, etc. Toute carte est une certaine interprétation de la réalité et non pas la réalité elle-même. Il ne faut donc pas conclure du déterminisme scientifiquement observé que la liberté n'existe pas ou qu'elle est, au contraire, absolue.

## *J. Intuition n°10 : La désincarnation progressive*

*« L'expérience de pensée de la désincarnation progressive prouve que je suis quelque chose d'essentiellement différent de tout objet physique situé dans l'espace-temps. »*

Cf. l'analyse de l'intuition n°4.

Critique :

On peut arguer contre cela que le simple fait qu'un scénario soit conceptuellement, voire métaphysiquement possible n'implique pas qu'il soit métaphysiquement vrai. On peut imaginer qu'en bougeant ses deux bras on va réussir à s'envoler. C'est sans doute conceptuellement et métaphysiquement possible, mais cela demeure faux – et physiquement impossible.

## *K. Intuition n°11 : Les zombies*

*« Je peux concevoir un être qui serait microphysiquement indiscernable de moi, mais qui n'aurait pas de conscience phénoménale (ce que les philosophes appellent un « zombie »). Cette dernière ne saurait donc se réduire à la matière. »*

Critique :

Cette expérience de pensée est l'inverse de la précédente et appelle les mêmes remarques.

De plus, on peut douter qu'il soit véritablement possible de concevoir, idéalement et positivement[100], un être sans conscience phénoménale. Comme le note Marcus :

> « On suppose que l'absence de conscience phénoménale des zombies est un fait subjectif les concernant. D'un autre côté, il n'y a aucun effet que cela fait d'être subjectivement absent. Donc il n'y a pas d'imagination du type "effet que cela fait" que nous pouvons utiliser pour arriver à cette possibilité »[101]

Pour Shoemaker[102], les zombies sont également inconcevables, mais pour une autre raison : nous avons une connaissance de nos propres expériences. Chez les zombies, les mêmes processus introspectifs les mènent à la conclusion qu'ils ont aussi des expériences,

---

[100] Sur la signification de ces termes, dans le contexte de la concevabilité de scénarios imaginaires, voir le chapitre 2 : « Du concevable au possible », de la partie 3 : « Identité personnelle et expériences de pensée », de cette thèse.

[101] MARCUS, Eric. "Why Zombies are Inconceivable".

[102] SHOEMAKER, Sidney. "Functionalism and Qualia".

mais cette conclusion est fausse. Par conséquent, si les zombies sont concevables, alors il est faux d'affirmer que nous connaissons nos expériences. En effet, puisque les zombies se trompent du tout au tout en pensant en avoir, quelle garantie avons-nous de ne pas nous tromper au sujet des nôtres ? Comment pouvons-nous seulement savoir que nous ne sommes pas des zombies nous-mêmes ?

Enfin, dans le chapitre 4 de cette partie : « De l'internalisme à l'externalisme », nous montrerons qu'il y a de bonnes raisons de supposer un lien constitutif entre action, perception et aspects phénoménaux de la conscience, si bien qu'un être qui se comporterait comme s'il était le siège de telle ou telle expérience serait nécessairement un être qui est le siège de cette expérience.

## *L. Intuition n°12 : Le néant subjectif inimaginable*

*« Je peux imaginer mon corps complètement détruit, dans la mort par exemple, mais je ne peux pas imaginer ma propre inexistence, en première personne. Je suis donc immortel, et différent de mon corps. »*

Cf. ce que nous avons déjà dit, en introduction de ce chapitre sur les concepts d'apparition et de disparition. Pour un développement plus approfondi, nous renvoyons notre lecteur à la partie 3 : « Identité personnelle et expériences de pensée », ainsi qu'aux annexes : « Identité personnelle et survie ».

## *M. Conclusion*

Les intuitions concernant l'identité personnelle sont très vivaces, toutefois, sur un plan strictement rationnel et/ou empirique, elles n'ont pas d'assise suffisamment solide. Chaque fois, ces intuitions doivent s'appuyer sur la prise en considération de ressentis subjectifs. Cela ne veut pas dire qu'elles sont fausses et absolument injustifiables. Au contraire, nous avons montré, par exemple dans la critique de l'intuition n° 9, comment la croyance au libre-arbitre peut être justifiée par une double considération :

L'*intuition* (qui nous faire croire, entre autres, au libre-arbitre) n'est, épistémiquement parlant, ni plus ni moins justifiée que la raison et l'expérience (qui nous fait en douter) ; c'est essentiellement des considérations pragmatiques d'utilité, de consensus intersubjectif et de renforcement hédonistique qui, dans les deux cas, tiennent lieu de justifications.

Du point de vue de la raison et de l'expérience (théorie néo-darwinienne de l'évolution) : l'*intuition*, mais aussi la raison et l'expérience, sont des processus biologiques dont la fonction est d'assurer la survie de l'individu et sa reproduction, et non pas de fournir des croyances vraies. Si notre confiance envers la raison et l'expérience comporte une dimension épistémique, alors cela fournit une justification indirecte à la confiance que nous pouvons avoir envers la valeur épistémique de notre *intuition*. En d'autres termes, la

valeur épistémique de la raison et de l'expérience garantit la valeur épistémique de l'*intuition*[103].

Pour les mêmes motifs, on peut justifier la croyance en une identité personnelle, un *moi* et/ou un *sujet*, ontologiquement distinct de la matière, car cette croyance repose sur une intuition très forte, très répandue, et sans conteste utile, puisqu'elle répond à un besoin, non moins universel, de survivance par-delà la mort. Saint Thomas d'Aquin dit quelque chose de très semblable, lorsque, dans sa *Somme Théologique*[104], il énonce :

> « L'âme humaine ne peut […] être corruptible. On peut trouver une preuve de cette incorruptibilité dans cette vérité générale : tout être désire naturellement exister, sous le mode qui lui convient. Chez les êtres dotés de connaissance, le désir est proportionné au mode de connaître. Le sens ne connaît l'être que dans une étendue et une durée concrètes, mais l'intelligence le connaît absolument, et par référence à n'importe quel temps. Aussi, tout être doté d'intelligence désire-t-il naturellement exister toujours. Mais un désir naturel ne peut être vain. Toute substance intelligente est donc incorruptible. »

---

103 Pour autant, cela va sans dire, qu'il n'y ait pas contradiction entre les croyances intuitives et les croyances empirico-rationnelles, et que, notamment, l'intuition ne nous amène pas à nier la théorie néo-darwinienne de l'évolution qui en garantit la valeur épistémique. Il faut souligner toutefois que s'il y a contradiction entre nos croyances intuitives et empirico-rationnelles, on peut considérer que cette contradiction réside davantage dans nos représentations de la réalité que dans la réalité elle-même ; ainsi, en physique, les contradictions entre la théorie quantique et la théorie de la relativité ou entre le modèle ondulatoire et le modèle corpusculaire de la lumière, même si elles devaient ne jamais être résolues, n'empêcherait pas de considérer ces divers modèles et théories comme justes et valables jusqu'à un certain point, car métaphysiquement, il y a toujours moyen de rendre compatible ce qui de prime abord paraît incompatible : par exemple, la lumière pourrait être juste ce qu'elle nous apparaît être dans les diverses circonstances au cours desquelles nous l'étudions : une entité ayant des propriétés ondulatoires et corpusculaires, même si cela nous paraît difficile à se représenter en imagination, surtout compte tenu des préjugés newtoniens que nous avons sur ce que doit être la matière.

104 D'AQUIN, Saint Thomas, *Somme Théologique*, Partie 1, Question 75, Article 6.

Une autre manière, enfin, de justifier nos intuitions antiréductionnistes sur l'identité personnelle consiste à montrer en quoi un parallèle peut être établi entre elles et nos intuitions sur l'existence d'un monde extérieur, indépendant de notre appréhension de ce dernier, et que nos meilleures théories scientifiques tendent idéalement et asymptotiquement à décrire tel qu'il se présente réellement. Comme le note Uwe Meixner[105] :

> « Mon objectif dans cet article n'est pas de critiquer l'Inférence à la Meilleure Explication, laquelle en effet est et a été d'une importance primordiale dans l'aventure humaine. J'aimerais seulement souligner le problème qui est que la plupart des réalistes sont injustement sélectifs dans l'emploi de cette forme d'argumentation.
>
> Ils accueillent bien son emploi dans l'établissement de certaines formes de réalisme, alors qu'ils le rejettent dans toute tentative d'établir d'autres formes de réalisme, *bien que* ces dernières formes de réalisme sont *compatibles* avec la première. Au contraire, ils affirment même généralement les négations des réalismes en question. Cela sème le doute quant à la rationalité métaphysique de la plupart des réalistes : apparemment ils ont dès le départ, *concernant ce qui n'existe pas*, certains préjugés, qu'aucune inférence à la meilleure explication, ou autre argument, ne permet d'ébranler.
>
> Bien sûr, la plupart des réalistes nieront ce fardeau ; ils brandiront le Rasoir d'Occam […].
>
> Malheureusement, le Rasoir d'Occam peut être utilisé pour éliminer *tout* ce que l'Inférence à la Meilleure Explication peut introduire. Pourquoi alors le Rasoir d'Occam est appliqué dans certains cas, et n'est pas appliqué dans d'autres où il pourrait aussi être appliqué ? »

Or, s'il y a bien un cas où une option réaliste peut être défendue avec la même force que pour les théories scientifiques ou le sens commun, c'est bien dans le cas du « moi », de la personne. Le réalisme du sujet, n'est que le juste symétrique du réalisme de l'objet, lequel

---

[105] MEIXNER, Uwe, "On Some Realisms Most Realists Don't Like". [conférence]

est admis communément, par intuition préthéorique, même si on cherche à le justifier *a posteriori* par une inférence à la meilleure explication.

De même que l'existence d'une réalité extérieure constituerait la meilleure explication au succès de nos théories scientifiques, au caractère parfois imprévu de nos observations du monde, ainsi qu'à la résistance de la matière à notre action, l'existence d'une réalité intérieure pourrait constituer la meilleure explication au succès de nos théories morales, à notre liberté et à nos intuitions antiréductionnistes. De plus, la ressemblance et la cohérence de nos perceptions, souvent citées en appui de l'existence du monde extérieur, pourraient tout aussi bien s'expliquer par l'existence d'un monde intérieur. Pour David Hume[106], notre croyance en l'identité des objets et en notre propre identité de sujet s'enracine dans la perception, par le jeu de l'imagination et de la mémoire :

> « [C]'est une fausse opinion que de penser que l'un quelconque de nos objets ou l'une quelconque de nos perceptions est identiquement le même après une interruption ; et, par conséquent, l'opinion de leur identité ne peut jamais naître de la raison mais doit provenir de l'imagination. L'imagination est entraînée vers une telle opinion seulement au moyen de la ressemblance de certaines perceptions puisque ce sont seulement nos perceptions ressemblantes que nous avons une propension à supposer identiques. Cette propension à conférer une identité à nos perceptions ressemblantes produit la fiction d'une existence continue puisque cette fiction, aussi bien que l'identité, est réellement fausse, comme le reconnaissent tous les philosophes, et n'a d'autre effet que de remédier à l'interruption de nos perceptions, seule circonstance contraire à leur identité. En dernier lieu, cette propension cause la croyance au moyen des impressions présentes de la mémoire puisque, sans le souvenir des perceptions précédentes, il est clair que nous n'aurions jamais aucune croyance à l'existence continue des corps. Ainsi, en examinant toutes ces parties, nous trouvons que chacune d'elles

[106] HUME, David. *Traité de la Nature Humaine*, partie IV, section 6 : « De l'identité personnelle ».

> repose sur les plus fortes preuves, et qu'elles forment toutes ensemble un système cohérent parfaitement convaincant. Une forte propension ou inclination seule, sans aucune impression présente, causera parfois une croyance ou une opinion. Combien davantage quand elle est aidée par cette circonstance ! »

Plus loin, l'auteur s'emploie à une comparaison entre le problème de l'identité personnelle et celui de l'identité des objets :

> « Nous passons maintenant à l'explication de la nature de l'*identité personnelle*, qui est devenue une question si importante en philosophie, surtout ces dernières années en *Angleterr*e où toutes les sciences les plus abstruses sont étudiées avec une ardeur et une application particulières. Et ici, il est évident que la même méthode de raisonnement doit être suivie, celle qui nous a expliqué avec tant de succès l'identité des plantes, des animaux, des bateaux, des maisons et de toutes les productions composées et changeantes, soit de l'art, soit de la nature. L'identité que nous attribuons à l'esprit de l'homme est une identité fictive du même genre que celle que nous attribuons aux corps végétaux et animaux. Elle ne peut donc avoir une origine différente. Elle doit procéder d'une semblable opération de l'imagination sur des objets semblables. »

Enfin, David Hume enracine l'illusion de l'identité personnelle dans les relations de *ressemblance* et de *causalité* entre les perceptions. Puis il nous montre le rôle essentiel de la mémoire dans ce processus :

> « Pour commencer par la *ressemblance*, supposez que nous puissions voir clairement à l'intérieur d'autrui et observer cette succession de perceptions qui constitue son esprit ou son principe pensant, et supposez qu'il conserve encore la mémoire d'une partie considérable des perceptions passées. Il est évident que rien ne saurait contribuer davantage à accorder à cette succession une relation au milieu de toutes ses variations. En effet, qu'est-ce que la mémoire, sinon une faculté par laquelle nous éveillons les images des perceptions passées ? Et, comme une image ressemble nécessairement à son objet, le fait de placer ces

> perceptions ressemblantes dans la chaîne de la pensée ne doit-il pas conduire l'imagination plus facilement d'un maillon à un autre et faire que le tout paraisse semblable à la persistance d'un objet ? D'ailleurs, sur ce point, la mémoire, non seulement découvre l'identité, mais contribue aussi à sa production en produisant la relation de ressemblance entre les perceptions. Le cas est le même, que nous nous considérions nous-mêmes ou que nous considérions autrui.
>
> Pour ce qui est de la *causalité*, nous pouvons observer que la véritable idée de l'esprit humain est de le considérer comme un système de différentes perceptions ou de différentes existences qui sont enchaînées les unes aux autres par la relation de cause à effet, et qui se produisent, se détruisent, s'influencent et se modifient les unes les autres. Nos impressions donnent naissance à leurs idées correspondantes, et les idées, à leur tour, produisent d'autres impressions. Une pensée en chasse une autre, et attire une troisième par laquelle elle est chassée à son tour. A cet égard, je ne peux comparer plus proprement l'âme qu'à une république, un Etat, dans lequel les différents membres sont unis par les liens réciproques de gouvernement et de subordination, donnent naissance à d'autres personnes qui reproduisent la même république dans les changements incessants de ses parties. Et, tout comme la même république particulière peut changer, non seulement ses membres, mais aussi ses lois et ses constitutions, de manière semblable, la même personne peut changer de caractère et de disposition, aussi bien que d'impressions et d'idées, sans perdre son identité. Quelques changements qu'elle subisse, ses différentes parties sont toujours en connexion par la relation de causalité. Et, à ce point de vue, l'identité qui concerne nos passions sert à corroborer celle qui concerne notre imagination, en faisant que nos perceptions distantes s'influencent les unes les autres, et en nous donnant un intérêt présent à nos douleurs et à nos plaisirs passés ou futurs. »

Il est intéressant de noter, contre l'opinion commune qui fait de David Hume un parfait antiréaliste en matière d'identité personnelle, que celui-ci évoque la possibilité selon laquelle l'*impression* de l'identité personnelle pourrait être une *perception véridique*, et non une *hallucination* ou une *illusion* (bien sûr, elle pourrait l'être de temps en temps, à l'instar de nos

impressions des objets, qui sont parfois des hallucinations ou des illusions, mais qui sont généralement des perceptions véridiques). Ainsi, David Hume écrit :

> « Comme la mémoire seule nous fait connaître la persistance et l'étendue de cette succession de perceptions, elle doit être considérée, pour cette raison principalement, comme la source de l'identité personnelle. Si nous n'avions pas de mémoire, nous n'aurions jamais aucune notion de causalité, ni par conséquent de cette chaîne de causes et d'effets qui constitue notre moi, notre personne. Mais une fois que nous avons acquis cette notion de causalité par la mémoire, nous pouvons étendre la même chaîne de causes, et par conséquent l'idée de notre personne, au-delà de notre mémoire et nous pouvons englober les moments, les circonstances et les actions que nous avons complètement oubliés mais dont nous supposons en général l'existence. En effet, peu nombreuses sont les actions passées dont nous ayons quelque mémoire. Qui peut me dire, par exemple, quelles furent ses pensées et ses actions le 1er janvier 1715, le 11 mars 1719 et le 3 août 1733 ? Ou affirmera-t-on, parce qu'on a entièrement oublié les incidents de ces jours, que le moi présent n'est pas la même personne que le moi de cette époque, et, de cette façon, mettra-t-on sens dessus dessous les notions les mieux établies d'identité personnelle ? De ce point de vue, donc, la mémoire ne *produit* pas tant qu'elle ne *découvre* l'identité personnelle, en nous montrant la relation de cause à effet entre nos différentes perceptions. Il incombera à ceux qui affirment que la mémoire produit entièrement notre identité personnelle de donner la raison pour laquelle nous pouvons ainsi étendre notre identité personnelle au-delà de notre mémoire. »

Cela rejoint sensiblement l'avis de Leibniz pour qui :

> « Un être immatériel ou un esprit ne peut être dépouillé de toute perception de son existence passée. Il lui reste des impressions de tout ce qui lui est autrefois arrivé et il a même des pressentiments de tout ce qui lui arrivera ; mais ces sentiments sont le plus souvent trop petits pour pouvoir être distingués et pour qu'on s'en aperçoive, quoiqu'ils puissent peut-être se développer un jour. Cette continuation et liaison de perceptions fait le même individu réellement, mais les aperceptions

> (c'est-à-dire lorsque l'on s'aperçoit des sentiments passés) prouvent encore une identité morale, et font paraître l'identité réelle. La préexistence des âmes ne nous paraît pas par nos perceptions, mais si elle était véritable, elle pourrait se faire connaître un jour. Ainsi il n'est point raisonnable que la restitution du souvenir devienne à jamais impossible, les perceptions insensibles [N.B. : Comprendre les sensations et sentiments inconscients, non perçus, autrement dit ce que Leibniz appelle aussi : « petites perceptions »] (dont j'ai fait voir l'usage en tant d'autres occasions importantes) servant encore ici à en garder les semences. »[107]

Par le fait, et malgré la déconstruction rationnelle que nous avons effectuée des intuitions antiréductionnistes de l'identité personnelle, nous sommes quand même fondés, non pas en pure raison, mais pragmatiquement, à y adhérer. Mais, puisque c'est également pragmatiquement que nous sommes fondés à accorder confiance à notre raison, cette assise pragmatique à nos intuitions antiréductionnistes sur l'identité personnelle n'a aucune raison d'être épistémiquement dépréciée. Selon Roger Pouivet, on ne peut pas vraiment douter que le monde extérieur existe mais simplement faire semblant d'en douter.

Dans *Epistémologie de la croyance religieuse*, résumant des propos de Wittgenstein présents dans son livre *De la certitude*, Roger Pouivet écrit[108] :

> « [I]l y a au moins cinq raisons de penser que la conception cartésienne du doute est indéfendable. (1) Les doutes supposent des raisons de douter. (2) Les doutes doivent faire une différence dans la pratique ; ils ne peuvent être purement spéculatifs, sauf à être purement verbaux. (3) Les doutes présupposent la maîtrise de jeux de langage (et

---

[107] LEIBNIZ, Gottfried Wilhelm, *Les Nouveaux Essais sur l'Entendement Humain*, Livre II, Chapitre XXVII (dans le passage cité, Leibniz s'exprime à travers le personnage de Théophile)

[108] POUIVET, Roger, « Epistémologie de la croyance religieuse », *In :* BOURGEOIS-GIRONDE, Sacha, GNASSOUNOU, Bruno et POUIVET, Roger, (éd.), *Analyse et théologie : croyances religieuses et rationalité.*

> de tout ce qui socialement les accompagne). (4) Aucun doute ne peut porter sur tout à la fois (c'est simplement absurde). (5) Le doute présuppose donc la certitude (puisqu'il ne peut porter sur tout à la fois). »

Plus loin[109], Roger Pouivet cite Thomas Reid[110] :

> « Nous sommes nés sous la nécessité d'avoir confiance dans notre capacité de raisonnement et de jugement, et le plus grand sceptique ne peut maintenir une croyance réelle de leur caractère trompeur pendant un temps un peu long, car c'est faire violence à sa constitution. C'est comme un homme qui marche sur les mains, un tour que certains hommes peuvent faire à l'occasion, mais aucun homme ne peut faire ainsi un long voyage. Cessez d'admirer sa dextérité, et comme les autres hommes, il se remettra sur ses jambes. »

La même ligne d'argumentation peut nous convaincre, dans le cas de l'identité personnelle, de l'impossibilité, sinon discursive, de la nier, quand bien même nous ne pouvons argumenter rationnellement en sa faveur. Et plus, de la nécessité d'avoir confiance en le fait que nous sommes un moi irréductiblement subjectif. Cette confiance fondant notre pratique dans un nombre impressionnant de cas : responsabilité morale, vie sociale, intérêt pour notre futur, sentiments de culpabilité, de vengeance, d'indignation et de fierté, récompense et punition, amour et haine, mérite et démérite, etc.

---

[109] POUIVET, Roger, « Epistémologie de la croyance religieuse », *In :* BOURGEOIS-GIRONDE, Sacha, GNASSOUNOU, Bruno et POUIVET, Roger, (éd.), *Analyse et théologie : croyances religieuses et rationalité.*

[110] REID, Thomas, *Essais sur les pouvoirs intellectuels de l'homme*, Essai VI, chap. V, p. 447.

# III. Des critères aporétiques

## *A. Critère psychologique :*

### 1. Circularité :

Le théologien anglais Joseph Butler (1692-1752), promoteur d'une approche rationnelle (usage des probabilités par exemple) de l'apologétique chrétienne en vue de convaincre les déistes à revoir leur considération de la foi catholique romaine, est l'auteur d'un argument contre les vues de Locke au sujet de l'identité personnelle. Selon cet argument, présenté dans son ouvrage majeur : *The Analogy of Religion*[111] : Le concept de souvenir présuppose conceptuellement celui de l'identité personnelle (on ne peut dire se souvenir avoir perdu une bataille à Waterloo le 18 juin 1815 si on n'est pas Napoléon I$^{er}$), par conséquent, Locke ne peut pas utiliser le critère mnésique pour fonder l'identité numérique des personnes car cela entraîne un problème de circularité : on définit l'identité personnelle par le souvenir, mais le souvenir est défini par l'identité personnelle. Pour répondre à cette objection, des néo-lockéens tels Sydney Shoemaker[112] et Derek Parfit[113] élaboreront le concept de quasi-mémoire. D'après Shoemaker, dire qu'une personne A se rappelle un événement X signifie :

(1) que A possède – présentement – la connaissance de X ;

---

111 BUTLER, Joseph, *The Analogy of Religion*, Premier appendice : « Of Personal Identity ».

112 SHOEMAKER, Sidney. *The First-Person Perspective and Other Essays.*

113 PARFIT, Derek. *Reasons and Persons.*

(2) que la connaissance de A n'est pas fondée par induction ou basée sur le témoignage d'autres personnes ;

(3) que A a été témoin de X lorsqu'il s'est produit.

Le point 3, cependant, présuppose que nous disposions d'un critère de l'identité personnelle qui nous permette d'identifier la personne ayant été témoin de l'événement X comme la même que celle qui satisfait le point 1. Evidemment, ce critère d'identité ne pourrait pas être le fait que A se rappelle de X, puisque nous connaîtrons ce fait que si nous avons préalablement établi que l'identité se maintient.

La quasi-mémoire ne pose pas ce genre de problème. D'après Shoemaker, dire qu'une personne A se quasi-rappelle un événement X signifie seulement :

(1) que A possède – présentement – la connaissance de X ;

(2) que la connaissance de A n'est pas fondée par induction ou basée sur le témoignage d'autres personnes ;

Le point 3, constitutif du concept de mémoire est absent de celui de quasi-mémoire, et la quasi-mémoire peut donc servir à fonder l'identité personnelle à travers le temps, en échappant à l'objection butlerienne de circularité.

## 2. Non-transitivité

Dans ses *Essais sur les Facultés de l'Esprit humain*[114], Thomas Reid écrit :

> « Un militaire a reçu le fouet au collège pour avoir dévalisé un verger ; il a pris un drapeau à l'ennemi dans sa première campagne ; plus tard il est devenu général. Lorsqu'il a pris le drapeau, il se souvenait de l'aventure du collège ; lorsqu'il est devenu général, il se souvenait d'avoir pris le drapeau, mais il ne se souvenait plus d'avoir reçu le fouet à l'école.
>
> Dans les principes de Locke, celui qui a été fouetté à l'école est le même qui a pris le drapeau ; et celui qui a pris le drapeau est le même que celui qui est devenu général. Il s'ensuit rigoureusement que le général est le même que l'enfant puni au collège. Mais la conscience du général ne remontant point jusqu'aux coups de fouet qu'il a reçus, il suit de la doctrine de Locke que l'enfant et lui ne sauraient être la même personne. Donc le général est, et en même temps n'est pas la personne qui a reçu le fouet à l'école. »

Le raisonnement est implacable. Avec celui-ci à l'esprit, comment pourrait-on encore soutenir une thèse lockéenne de l'identité personnelle, ou même une thèse post-lockéenne améliorée aux quasi-mémoires ? L'identité numérique est transitive, le (quasi-)souvenir – et donc le critère mnésique, lockéen ou post-lockéen de l'identité personnelle – n'y est pas, car la capacité à se remémorer un certain épisode passé de notre vie n'implique nullement que l'on puisse aussi se remémorer des épisodes encore plus anciens dont nous nous rappelions alors. Le problème est général et dépasse le simple cadre des critères psychologiques de l'identité personnelle. Il surgit dès lors que l'on tente d'appliquer les catégories rigides de l'identité numérique sur le flot mouvant de l'existence, de la vie. Si on tolère le fondement d'une identité numérique parfaite, laquelle ne souffre aucun degré, et est affaire de tout-ou-

---

114 REID, Thomas. *Essais sur les facultés de l'esprit humain*, Essai Trois : « De la Mémoire », Chapitre 6 : « De l'Identité Personnelle, selon Locke »

rien, sur une identité qualitative approximative, alors de proche en proche, l'écart se creuse et le paradoxe jaillit.

Pour y répondre, il faut remanier nos concepts. Derek Parfit[115] nous propose, à ce titre, d'introduire ceux de connexité (*connectedness*) et de continuité (*continuity*).

Il les définit ainsi :

- **Connexité :** C'est le maintien de relations psychologiques directes (quasi-souvenir, cause à effet : quasi-intention et sa quasi-action par ex., etc.). C'est une relation non-transitive.

- **Continuité :** C'est une chaîne ininterrompue de relations psychologiques directes fortes se recouvrant partiellement.

Concevoir la connexité est aisé. Il s'agit grosso modo d'un lien de cause à effet direct entre deux états successifs. On peut aussi y inclure les relations d'identité qualitative. Si on a de nombreuses relations de connexité entre deux phases successives de la vie d'un individu on parlera de connexité forte et sinon de connexité lâche.

Pour bien visualiser en quoi consiste la continuité, vous pouvez, si vous voulez, vous imaginer des dominos alignés les uns derrière les autres. Vous poussez le premier et cela entraîne tous les autres à tomber. Le fait qu'un domino bouscule le suivant, c'est la connexité, le mouvement d'ensemble engendré, c'est la continuité. Si la connexité est

115 PARFIT, Derek. *Reasons and Persons*.

rompue quelque part (ex. : deux dominos successifs trop éloignés), le mouvement (la continuité) s'interrompt. Mais il n'est pas nécessaire que deux dominos soient proches l'un de l'autre pour que la continuité se transmette entre eux, il suffit qu'il y ait des dominos bien placés entre eux.

Appliquons ces concepts au problème soulevé par Thomas Reid. Il suffit de redéfinir le critère psychologique comme suit : A vivant en t1 et B vivant en t2 sont une seule et même personne si et seulement si, il existe une relation de (forte) connexité **ou** de continuité entre A et B.

A présent, nous nous retrouvons dans la configuration suivante :

Lorsque le militaire a pris le drapeau, il se souvenait d'avoir été fouetté au collège. Il y a donc connexité entre l'enfant et le militaire. Ce sont une seule et même personne.

Lorsqu'il est devenu général, il se souvenait d'avoir pris le drapeau, mais il ne se souvenait plus d'avoir reçu le fouet à l'école. Il n'y a pas connexité entre l'enfant et le général mais il y a connexité entre le militaire et le général. Ce sont une seule et même personne.

Celui qui a été fouetté à l'école est le même qui a pris le drapeau ; et celui qui a pris le drapeau est le même que celui qui est devenu général. Il s'ensuit rigoureusement que le général est le même que l'enfant puni au collège. Et puisqu'il y a connexité entre l'enfant et le militaire et entre le militaire et le général, il y a continuité entre l'enfant et le général. Il s'ensuit qu'il s'agit là d'une seule et même personne. Nous retrouvons donc bien, par

l'extension à la continuité des fondements de l'identité personnelle ce que la transitivité de l'identité numérique nous force à admettre.

### 3. Duplication :

Nous avons vu, précédemment, quelle objection de circularité pouvait être portée au critère mnésique, la mémoire présupposant d'avoir été le témoin de l'événement dont on se souvient, ce qui implique le maintien de l'identité personnelle entre ces deux moments. Le concept de quasi-mémoire, élaboré par Sidney Shoemaker puis par Derek Parfit, en faisant l'économie de cette présomption d'identité personnelle, peut servir de nouveau critère et échappe à la circularité. Malheureusement, le concept de quasi-mémoire va rendre possible une situation dans laquelle plusieurs personnes A, B, C, D, etc. peuvent se rappeler un même événement X, voire le même grand nombre d'événements X, Y, Z, etc. Par exemple, plusieurs personnes peuvent assister à une même scène en même temps et donc se souvenir du même événement. Deux frères siamois assisterons, en règle générale, de part leur condition particulière, aux mêmes événements extérieurs en même temps et pourront s'en rappeler. Stéphane Chauvier[116], autre néo-lockéen, propose une solution : seule la remémoration de *certains événements* et *sous un certain mode* est pertinente dans la constitution de l'identité personnelle. Il s'agit de la remémoration, *en première personne*, du *noyau plus ou moins permanent de croyances – fondées ou non, vraies ou fausses – que nous avons au sujet de nous-mêmes* qui est importante. Autrement dit, ce sont des propositions du type « je suis timide », « je m'appelle Mikaël », « je suis agnostique », « je suis de sensibilité centre-gauche », « j'aime les fraises », « je suis quelqu'un de sympathique », etc. qui vont fonder la trame de notre identité personnelle. Notons que cette remémoration ne se présente pas généralement

[116] CHAUVIER, Stéphane. *Qu'est-ce qu'une personne ?*

comme un souvenir (on ne dira pas : « je me rappelle que je suis timide » mais simplement « je suis timide »), et qu'on y intègre aussi, souvent, le point de vue d'autrui (la croyance liée à une proposition telle que « je suis quelqu'un de sympathique » est fortement liée, en général, au retour agréable et positif de la plupart des gens à notre égard). Bien entendu, il y a de nombreuses personnes différentes qui peuvent penser être timides, par exemple, mais il y en a probablement moins qui peuvent penser être à la fois timides et aimer les fraises, etc. Au final, on parviendra, à isoler au plus une personne susceptible d'exprimer l'ensemble de ces pensées égologiques. Du moins en pratique. Et c'est là, à notre sens, que la conception de Stéphane Chauvier, pour élaborée soit-elle, rencontre une limite. En théorie, rien n'interdirait à deux hypostases humaines numériquement distinctes d'incarner exactement le même ensemble de pensées égologiques permanentes. Bien sûr, la qualité de l'hypostase conditionne en grande partie les pensées égologiques. Un cul-de-jatte pourrait difficilement exprimer une pensée égologique du style « j'aime le jogging ». Mais deux hypostases peuvent être numériquement distinctes (par exemple, elles n'occupent pas le même espace au même moment et sont constitués de deux fractions de matière différentes) tout en étant qualitativement semblables.

Ce genre de situation a été popularisé par des expériences de pensée mettant en scène des scénarios de science-fiction impliquant l'existence de machines pour téléporter les personnes ou de greffes de cerveaux. Les philosophes contemporains David Wiggins, et surtout Derek Parfit sont très friands de ce genre d'argumentation.

Derek Parfit[117], par exemple, imagine des cas comme celui-ci : Une personne entre dans un téléporteur qui doit la transférer sur Mars. La machine scanne sa structure

[117] Parfit, Derek. *Reasons and Persons*.

moléculaire et envoie par ondes radios l'information à une machine similaire présente sur Mars. Ensuite, le corps de cette personne est désintégré. Finalement, le téléporteur présent sur Mars, à partir des informations émises de la Terre, reconstitue un individu indiscernable, à la molécule près, de l'individu de départ. Question : Cet individu est-il la personne de départ ou un simple Répliquant ? Si l'on s'accorde sur le fait que la conscience de soi de l'individu sur Mars contient les mêmes pensées égologiques que celle de l'individu qui était sur Terre, et si cela doit suffire comme critère de l'identité personnelle, alors oui, l'individu sur Mars est bien la personne originale et non un simple Répliquant. Et cela, quand bien même l'identité numérique de l'hypostase matérielle ne se conserve pas (les éléments constituant originalement le corps de l'individu de départ ne sont pas envoyés sur Mars, seule l'information relative à sa structure est envoyée, de manière à former, sur Mars et à partir des éléments présents sur place, un nouveau corps).

Mais imaginons cependant que, suite à un bug du téléporteur, l'information recueillie grâce au scanning de l'individu de départ soit envoyée, non pas seulement sur Mars, mais également sur Vénus. Et supposons que là, une autre machine captant le message reconstitue un individu moléculairement indiscernable de l'individu de départ, pendant qu'au même moment, un autre individu, également en tout point semblable à l'individu de départ, soit généré sur la planète Mars. Nous avons donc un individu qui entre, sur Terre, dans le téléporteur, et nous nous retrouvons avec deux individus, à l'arrivée, qui sont qualitativement identiques entre eux et à l'individu de départ. Un de ces individus se retrouve sur Mars, et l'autre sur Vénus. Ces deux individus sont-ils la personne de départ ou de simples Répliquants ? Les deux peuvent-ils être la personne de départ ? Si un seul au plus peut être la personne de départ, lequel est-ce et pourquoi ?

Si on appelle A la personne de départ, B la personne apparaissant sur Vénus et C la personne apparaissant sur Mars, on peut formaliser le problème comme suit :

En t2 on a deux personnes, B et C, qui sont psychologiquement en continuité avec la personne A qui existait en t1.

D'après le critère psychologique :

- B est en continuité avec A ⇔ A = (numériquement) B

- C est en continuité avec A ⇔ A = (numériquement) C

- L'identité numérique est transitive : B = (numériquement) C

Problème : B et C semblent être deux personnes différentes. En effet, elles peuvent ignorer totalement l'existence l'une de l'autre, évoluer très différemment, se rencontrer au bout de plusieurs dizaines d'années et ne pas se reconnaître l'une l'autre comme étant la même personne.

Pour résoudre ce problème, une série d'options philosophiques se présentent :

1. Assumer, au moins partiellement, la logique de l'identité personnelle, c'est-à-dire admettre que A = B et que A = C. On peut le faire d'au moins deux façons :

1.1. Considérer que A = B et que A = C et accepter la conclusion logique selon laquelle B = C : C'est-à-dire qu'il faudrait admettre, dans notre ontologie, la possibilité pour deux êtres n'occupant pas ou pas exactement le même espace au même moment d'être numériquement identiques.

1.2. Considérer que A = B et que A = C et refuser la conséquence logique suivant laquelle B = C : C'est-à-dire qu'il faudrait remettre en question la transitivité de l'identité.

2. Ne pas assumer la logique de l'identité personnelle et suggérer que ce cas constitue une exception. On peut le faire au moins de deux façons :

2.1. Considérer que A = B ou bien A = C : C'est-à-dire que des deux personnes B et C, au moins une et au plus une, sera la personne A de départ. L'autre est un « imposteur » qui s'ignore, un Répliquant.

2.2. Considérer que A ≠ B et A ≠ C : C'est-à-dire que B et C sont des « imposteurs » qui s'ignorent, ce sont tous deux des Répliquants.

3. Ne pas assumer la logique de l'identité personnelle et suggérer une reconceptualisation « exotique » de la notion de personne :

3.1. Considérer, en dépit des apparences, que B et C sont une même personne (A) existant à travers deux individus. Cette possibilité doit être distinguée de la première. Il ne s'agit pas de dire que B est

pleinement A et que C est pleinement A, mais que A est, après la duplication, un être formé de deux parties spatiales disjointes : B et C, qui sont donc à 50% A chacune.

3.2. Considérer, en dépit des apparences, que B et C sont alternativement dans le temps, la même personne (A), puis une autre personne. B et C sont donc 50% A chacune, mais en moyenne sur une certaine durée de temps, et non pas à chaque instant.

3.3. Considérer la thèse de l'occupation multiple[118] : L'individu A, en dépit des apparences, n'est pas une seule personne, mais deux (B et C) qui finissent par mener une existence séparée.

Après avoir exposé ces options, passons à présent à leur critique.

**a) Assumer, au moins partiellement, la logique de l'identité personnelle, c'est-à-dire admettre que A = B et que A = C.**

Le problème avec ces solutions c'est qu'elles nous forcent à admettre des idées qui ne vont pas du tout de soi, qui sont très lourdes et problématiques d'un point de vue ontologique ou logique. Qu'est-ce que cela voudrait dire, pour *deux* êtres n'occupant pas (tout à fait) le même espace au même moment, d'être *un* seul et même être ? N'y a-t-il pas déjà contradiction à dire que *deux* êtres sont *un* seul être ? Quand on dit que Phosphorus et

[118] Cf. PERRY, John. *Can the Self Divide?* Et : LEWIS, David Kellogg. "Survival and Identity". *In:* RORTY, Amélie Oksenberg. *The Identities of Persons*.

Hesperus sont un seul et même astre, il y a bien deux noms, et deux modes d'accès, mais ces deux noms et ces deux modes d'accès se réfèrent à une seule réalité. Là, il semblerait que l'on veuille parler de deux réalités, non pas qui se réfèreraient à une seule réalité, car se serait encore acceptable – en ce sens, Mars et la Terre sont deux réalités qui se réfèrent au Soleil, en tant qu'elles sont des planètes du système solaire – mais de deux réalités qui seraient une seule réalité. Là où on verrait deux choses apparemment distinctes et indépendantes, on nous demanderait d'admettre qu'il n'y a en fait qu'une seule chose. Bref, on nous demanderait d'admettre que de temps en temps, il arrive que 1 + 1 = 1 ! Si encore ces deux réalités apparentes évoluaient de concert, et que d'en modifier une modifiait l'autre instantanément, on pourrait admettre qu'elles ne forment qu'une seule réalité. Après tout, l'espace est peut-être relatif, la physique contemporaine tend à nous le faire voir ainsi, et peut-être qu'un objet pourrait réellement se trouver à deux localisations à la fois. C'est semble-t-il ainsi qu'il convient d'interpréter l'expérience d'Aspect [119] par exemple. Mais il est clair que ce n'est pas ce que l'on observerait dans le cas de B et C. Tout se passerait comme s'ils étaient bien numériquement distincts, aussi, on ne voit pas bien pourquoi on devrait les considérer comme numériquement identiques, si ce n'est pour résoudre le paradoxe de la duplication ! Mais cela serait résoudre un mystère en invoquant un mystère encore bien plus grand… Et quand on nous proposerait de remettre en question le principe logique de transitivité, on serait dans un cas de figure assez similaire. Quelle raison – au sens large de « motif » – aurions-nous à remettre en question ce pilier de la rationalité, sans lequel, des tas d'autres problèmes surgissent ? Car en effet, sans transitivité de l'identité, alors que Paul soit plus grand que Pierre et que Pierre soit plus grand que Jacques pourrait être compatible

[119] ASPECT, Alain & al.. "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers". Et : ASPECT, Alain & al. "Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities".

avec le fait que Paul est plus petit que Jacques. Peut-on se faire une représentation d'une telle situation ? Est-elle concevable ? Imaginable ? Il semble bien que non. Là encore, on tente de résoudre un mystère par un mystère encore plus grand, sans donner de motifs indépendants d'accepter ce second mystère.

### b) Ne pas assumer la logique de l'identité personnelle et suggérer que ce cas constitue une exception.

#### *(1) Considérer que A = B ou bien A = C :*

Tout le problème de cette option consiste à rendre compte du fait que concrètement, une des personnes va être la personne d'origine, mais pas l'autre. Mais laquelle sera-ce ? Et surtout, qu'est-ce qui va bien pouvoir justifier du fait que cela sera une des deux personnes qui sera la personne d'origine et pas l'autre ? Dans *Philosophical Explanations*, Robert Nozick, un philosophe contemporain malheureusement décédé depuis peu, et bien connu aussi pour ses écrits en philosophie politique, a émis la thèse, connue sous le nom de « thèse du meilleur candidat » ou « thèse du continuateur le plus proche »[120], selon laquelle, de B ou de C, ce sera celui qui satisfera au mieux le critère – celui défendu par ailleurs dans les cas ordinaires – de l'identité personnelle, qui sera numériquement identique à la personne A. Ainsi, de B – le « vénusien » – et de C – le « martien » – la thèse de Robert Nozick nous conduirait possiblement à affirmer que c'est ce dernier qui est authentiquement A, étant

[120] Ce n'est pas un nouveau critère, à côté des critères biologiques et psychologiques, mais une approche plus générale concernant l'emploi des critères existants, quelque soit celui que nous préférons. Cf. NOZICK, Robert, *Philosophical Explanations*.

donné l'intention première de A de se rendre sur Mars. Néanmoins, cette thèse souffre d'une faiblesse aisément reconnue par Nozick. Et si absolument rien ne permet de départager les différents candidats, que se passe-t-il ? Supposons que A n'ait pas souhaité se rendre sur Mars, mais sur la Lune, alors aucune des destinations finales ne serait en accord avec son intention de départ. Qui, de B ou de C serait alors numériquement la même personne que A ? De plus, que doit-on entendre par « le plus proche » ? Par exemple, si B existe avant C mais que B est psychologiquement plus proche de A que C, peut-on vraiment dire de B (ou de C) qu'il est un meilleur continuateur de A que C (ou B, respectivement) ? Face à de pareils cas, certains auteurs, tel Bruno Marchal[121], invoqueront une sorte d'*indéterminisme phénoménologique*. Ce serait alors une sorte de hasard ou de contingence qui « trancherait ». De nos yeux de philosophes, nous pourrions juste parier ou calculer une « probabilité de présence », à la manière de ce que l'on fait en physique quantique. Une option très proche invoquera un « fait manquant » (*further fact*) se superposant aux considérations sur la continuité psychologique et corporelle, et « expliquant » en dernier lieu la pérennité de l'identité. Un tel fait, pour remplir correctement son rôle, devrait présenter la propriété spéciale d'être atomique, non-reproductible, subjectif. Certains philosophes font le rapprochement entre ce fait hypothétique et l'hypothèse cartésienne de l'âme, mais ce n'est qu'une manière de l'interpréter. Ce « fait manquant » est à vrai dire assez peu crédible. Cela ressemble à une pétition de principe. Il n'y a rien de vérifiable qui puisse faire appui à cette thèse, et à la réflexion, elle paraît même inutile, ne réglant qu'en apparence seulement le problème posé par la duplication. En effet, on pourrait imaginer que ce fait puisse permuter d'un individu à l'autre toutes les 5 minutes sans que personne ne puisse s'en apercevoir… Le soulagement cognitif que cette hypothèse apporte est donc factice et provisoire. Elle ne fait

[121] MARCHAL, Bruno. *Calculabilité, Physique et Cognition*.

que repousser le problème un cran plus loin, et couvrir notre ignorance d'un peu de dignité en l'érigeant en thèse métaphysique…

Une autre option est envisageable. On pourrait argumenter que l'expérience de pensée proposée n'est tout simplement par possible théoriquement. Cela peut être, par exemple, à cause de l'une des raisons suivantes :

- On pourrait penser qu'il n'existe pas, et même qu'il ne peut pas exister, dans la nature, d'êtres parfaitement indiscernables, ou bien d'êtres différents suivant exactement le même degré d'un troisième. Par conséquent, de B et de C, il y en aurait nécessairement un des deux qui serait plus ressemblant à A que l'autre. Celui-là serait authentiquement A.

- On pourrait, suivant le même ordre d'idée, penser qu'il est impossible que deux événements soient exactement simultanés dans le temps. Par conséquent, de B et de C, il y en aurait nécessairement un qui serait créé avant l'autre. Le premier venu serait authentiquement A.

C'est difficile à croire. Cela serait peut-être envisageable sous l'hypothèse d'un espace-temps continu ou les variables de temps et d'espace pourraient prendre une infinité de valeurs. En ce cas, il serait virtuellement impossible d'obtenir *exactement* la situation décrite dans l'expérience de pensée. Mais, d'une part, on a de bonnes raisons de penser que

l'espace-temps est discontinu[122]. D'autre part, à supposer que l'espace-temps soit continu, il est difficile d'imaginer que des différences *qualitatives* si faibles entre B et C suffisent à fonder une identité *numérique*, laquelle présuppose une différence *radicale* entre B et C, ainsi qu'entre un des deux et A.

Cette difficulté est renforcée par le constat suivant : au cours d'une journée ordinaire, A, B et C pourront très bien être l'objet, chacun, de changements qualitatifs de bien plus grande amplitude (sommeil, veille, modifications d'humeur, oublis, apprentissages, etc.) que ceux qui – selon cette interprétation de notre expérience de pensée – sont supposés fonder une différence d'identité numérique. Pourtant, de tels changements ordinaires et quotidiens, que tout un chacun peut constater pour lui-même, ne semblent pas compromettre l'identité personnelle (quel que soit le critère que l'on utilise pour définir cette dernière). On ne « meurt » pas du simple fait d'avoir un bras arraché ou d'entrer soudain, après un profond calme, dans une grande colère…

Nozick, quant à lui, considère que si des cas pareils se présentaient, où B et C auraient exactement autant de raisons d'être considérés, chacun, comme le successeur de A,

---

[122] Avec la physique quantique. Voir aussi les articles de MAGNAN, Christian, *L'infini : ça n'existe pas !*, [en ligne] et *L'infini des cosmologistes : réalité ou imposture ?*, [en ligne], disponibles sur Internet, respectivement aux adresses suivantes : <http://www.lacosmo.com/infini.html> et <http://www.lacosmo.com/infini-encore.html>. Le premier article montre que l'infini est une notion mathématique qui n'a pas d'équivalent dans le monde physique. Soutenir que l'Univers est infini serait absurde car dépourvu de sens. Le second article développe les idées du précédent et montre que le concept d'infini est incohérent d'un point de vue physique et qu'il ouvre la porte à des dérives irrationnelles. Cette critique en règle de l'infini actuel a évidemment une portée quant à la question de la structure intime, continue ou discontinue, de l'Univers. En effet, supposer que l'Univers est continu revient à considérer que l'on peut le fractionner en quantités toujours plus petites, suivant un processus infini…

alors aucun des deux, de B ou de C, ne serait numériquement identique à A[123], ce qui nous amène à la critique de l'option suivante.

### *(2) Considérer que A ≠ B et A ≠ C :*

Repensons à la première expérience de pensée que nous avons imaginée. La personne, partant de la Terre, est téléportée sur Mars. On pourrait imaginer une situation similaire où la personne, partant de la Terre, est téléportée sur Vénus. Dans ces deux situations, l'identité numérique entre la personne de départ et celle d'arrivée ne pose pas de problème. Mais si on en croit l'option philosophique que nous nous proposons de critiquer à présent, il se passerait quelque chose de bien mystérieux, du moment où notre personne de départ serait téléportée à la fois sur Mars et sur Vénus. C'est comme si, malgré les millions de kilomètres de distance les séparant, l'existence de l'une allait avoir une répercussion sur l'existence de l'autre, l'empêchant d'être numériquement identique à la personne de départ, et par conséquent, cela aura des conséquences sur l'existence de la personne de départ également, puisque cette dernière n'existera plus. Autrement dit, s'il n'existait que B ou que C, on aurait respectivement : A = B et A = C, mais l'existence d'un second « candidat » à l'identité personnelle avec A « tuerait » A. C'est la théorie de l'absence de compétiteur, qui fut défendue tout d'abord par Wiggins [124].

#### (a) Une expérience de pensée :

[123] Pour Nozick, B (ou bien C) existant en t2 est une seule et la même personne que A existant en t1, si et seulement si, B (ou bien C, respectivement) est le meilleur continuateur de A, et qu'il n'y a pas d'autres continuateurs *ex aequo*.

[124] WIGGINS, David. *Sameness and Substance.*

Toute la bizarrerie de cette hypothèse ressortira mieux si nous la replaçons dans un scénario de science-fiction. Imaginez qu'une personne – appelons-là Adolf – vous a fait un très mauvais coup et vous ruminez de vengeance contre elle et projetez de la tuer. Malheureusement, cette personne est beaucoup plus forte que vous, entourée d'une équipe de protection personnelle et vous savez que de toute manière, même si vous réussissiez à l'abattre, les soupçons ne tarderont pas à se retourner contre vous. De plus, vous êtes scrupuleux et pensez à ses proches, qui ne vous ont rien fait, et qui seraient inconsolables de cette perte. Une idée géniale vous vient cependant à l'esprit. Vous apprenez que demain matin, votre ennemi juré doit se rendre, par téléportation, à un important congrès sur Pluton. Grâce à vos relations, vous parvenez à connaître l'heure précise de son départ. Or, il se trouve que vous avez chez vous un téléporteur, ainsi qu'une antenne apte à capter des signaux radios en provenance de téléporteurs. Votre plan, diabolique, est le suivant : lorsque votre ennemi juré se téléportera demain pour Pluton, vous capterez le signal radio émis et l'utiliserez pour recréer, avec votre téléporteur personnel, un deuxième exemplaire d'Adolf dans votre garage, au même moment où l'exemplaire plutonien apparaîtra. La présence de ce deuxième exemplaire d'Adolf aura pour répercussion de transformer le voyage du véritable Adolf en un voyage-suicide (même si personne ne s'apercevra de rien), puisque aucun des « Adolf » reconstitué par les téléporteurs ne sera l'authentique. Celui qui nous avait fait un mauvais coup et contre lequel nous nourrissions de la rancune, celui-là aura eu ce qu'il mérite, il aura tout bonnement cessé d'exister. Comme aurait dit Monsieur Cyclopède durant sa *Minute nécessaire* : Etonnant, non ?

**(b) Identité personnelle et survie :**

Pour Derek Parfit, il faut distinguer survie et identité personnelle (au sens de l'identité numérique), et donc ce n'est pas si étonnant. Pour Parfit, il semble que je peux survivre dans le futur, quand bien même personne dans ce futur n'est moi. Ce qui importe vraiment, ce n'est pas la conservation de l'identité personnelle, c'est seulement le maintien des relations de connexité et de continuité psychologiques au cours du temps. Celles-ci ne sont pas des relations nécessairement *« one-to-one »* mais peuvent très bien être des relations *« one-to-many »*. Donc Adolf – ou plutôt sa subjectivité, Adolf comme sujet singulier – ne sera pas tué, annihilé, par la reconstitution de deux exemplaires de lui-même, quand bien même aucun de ces exemplaires de lui-même n'est *numériquement* lui-même. Mais si Adolf a survécu, où se trouve-t-il ? Sur Pluton ou dans votre garage ? Il semble difficile d'échapper à la logique de l'identité personnelle et en même temps de tenir un discours cohérent sur la survie d'Adolf. Le mieux serait de dire qu'Adolf, bien que son identité personnelle n'ait pas été maintenue, survit à présent à travers deux individus : un se trouvant sur Pluton, l'autre dans votre garage. Mais quel effet cela fait-il de survivre à travers deux individus ? Si cela ne fait aucun effet particulier, alors pourquoi parler de survie ? Si cela fait un effet, quel est-il ? Manifestement, il n'y a personne qui se sent être à la fois à un congrès sur Pluton, insouciant, et enfermé dans votre garage, se demandant anxieusement ce qui s'est passé et ce qui va encore arriver… Les deux situations sont d'ailleurs logiquement incompatibles pour une seule personne. Admettons même qu'il existe une télépathie constante entre l'Adolf de Pluton et l'Adolf de votre garage, cela ne pourrait pas rester confiné dans l'intériorité d'une subjectivité. Cela aurait des conséquences observables. Par exemple, la copie plutonienne d'Adolf s'écrierait : « C'est incroyable, je suis ici avec vous sur Pluton, et en même temps, une partie de moi-même est prisonnière dans un garage sur Terre ! ». Elle ne pourrait donc pas assister à son congrès dans l'insouciance la plus pure. De plus, on se retrouverait dans la logique de l'identité personnelle que Parfit veut nous inviter à

abandonner. La copie plutonienne d'Adolf s'identifierait à la copie terrestre dans votre garage et en même temps à l'Adolf original avant d'entrer dans le téléporteur. Il semble peu plausible qu'elle dise quelque chose du genre : « C'est incroyable, je suis ici avec vous sur Pluton à penser mes pensées, percevoir mes perceptions, vouloir mes volitions, accomplir mes actions, et en même temps, je pense les pensées, je perçois les perceptions, je veux les volitions et j'accomplis les actions d'un type emprisonné dans un garage sur Terre, qui croit être moi et qui me ressemble comme deux gouttes d'eau ». Elle utilisera plus vraisemblablement le pronom « *je* » pour qualifier le deuxième Adolf, ou au minimum elle en réfèrera avec une expression du genre : « partie de *moi-même* ». Ce qui nous amène à la critique de l'option suivante.

### c) Ne pas assumer la logique de l'identité personnelle et suggérer une reconceptualisation « exotique » de la notion de personne :

#### *(1) Considérer, en dépit des apparences, que B et C sont une même personne (A) existant à travers deux individus.*

Comme nous l'avons déjà mentionné, il ne s'agit pas de dire que B est pleinement A et que C est pleinement A, mais que A est, après la duplication, un être formé de deux parties spatiales disjointes : B et C, qui sont donc à 50% A chacune. Adolf, après sa téléportation sur Pluton et dans votre garage, est une personne constituée de deux corps, deux têtes, quatre yeux, quatre oreilles, deux bouches, quatre bras et quatre jambes. Pourtant, Adolf n'expérimente pas cette double individualité. Tout se passe comme si on

avait en fait deux personnes chacune incarnée dans un individu différent et non pas une personne incarnée par deux individus différents. La partie plutonienne est consciente d'elle-même comme un tout indivisible et considère que la partie terrestre n'a aucune relation avec elle. La partie terrestre est consciente d'elle-même comme un tout indivisible et considère que la partie plutonienne n'a rien à voir avec elle. Les deux parties peuvent développer des personnalités très différentes au fil du temps, et s'ignorer complètement. Se croisant par hasard, elles ne se reconnaîtront pas l'une l'autre. Comment le comprendre ? Si Adolf n'a pas le sentiment d'être à deux endroits à la fois et de voir le monde à travers 4 yeux et 4 oreilles, alors on ne peut pas dire que Adolf-Pluton et Adolf-Terre forment une seule et même personne.

**(a) Une personne peut être deux individus à la fois, sans expérimenter cette double individualité :**

Pour comprendre comment cela peut être malgré tout possible, Derek Parfit nous suggère de réviser notre concept traditionnel de personne. Une personne peut être deux individus en même temps, même si elle n'expérimente pas cette double individualité ; pour la même raison qu'une personne peut être deux individus l'un après l'autre, même si elle n'expérimente pas cette double individualité. Imaginons que les jours pairs, Adolf ne vive qu'à travers Adolf-Pluton (pendant ce temps, supposons qu'Adolf-Terre soit animé par Rodolphe, un esprit très semblable à Adolf mais numériquement différent). Les jours impairs, Adolf ne vit qu'à travers Adolf-Terre (et pendant ce temps, Rodolphe vit à travers Adolf-Pluton). Mais en changeant de corps, supposons que chaque esprit perde entièrement la mémoire et la personnalité du corps précédent et embrasse entièrement la mémoire et la personnalité du corps suivant, alors, à tout moment du temps, tout se

passerait, pour Adolf ou Rodolphe, comme si ils étaient constamment liés au même corps, à la même personnalité, à la même mémoire. L'interversion quotidienne ne serait pas expérimentée. Mais peut-on soutenir une telle idée et continuer de défendre un critère psychologique, sinon de l'identité personnelle (Derek Parfit y a renoncé) mais au moins de la survie ? Derek Parfit, rappelons-le, définit la survie comme le maintien de relations de connexité et de continuité psychologiques au cours du temps, selon un mode « *one-to-one* » ou « *one-to-many* ». Dans le cas « *one-to-many* », Adolf survit à travers Adolf-Pluton et Adolf-Terre. Peut-on dire que Adolf-Pluton survit à travers Adolf-Terre ou que Adolf-Terre survit à travers Adolf-Pluton ? Outre que la définition parfitienne de la survie implique la succession temporelle, on voit clairement que Adolf-Terre et Adolf-Pluton ne sont pas en relation de connexité ou de continuité psychologiques. Même s'ils se rappellent un passé commun, leur mémoire, à partir de l'événement de duplication, se construit séparément et sans contact, et leur personnalité peut prendre des virages très différents. S'il y a bien transitivité de la relation d'identité, la relation « X survit en tant que Y » n'est, en revanche, pas transitive, selon Parfit, ou plutôt pas nécessairement transitive car elle peut cependant l'être parfois. Considérez la moitié droite de votre corps. Nul doute que cette moitié droite vit aussi à travers la moitié gauche de votre corps. Considérez la moitié gauche de votre corps. Nul doute que cette moitié gauche vit aussi à travers la moitié droite de votre corps. Imaginons que vous alliez à la guerre. Vous revenez amputé de votre jambe et bras droits. Certes, vous aurez survécu, mais vous considérerez qu'il y a eu des séquelles. Vous avez, en grande partie, survécu mais vous êtes un peu mort, votre jambe et bras droits sont morts. Chacune de vos moitiés est une partie essentielle à votre survie pleine et entière. L'absence de l'une est un événement catastrophique pour l'autre. Considérons à présent ceci : seul Adolf-Pluton aurait pu exister, et cela n'aurait aucunement compromis la survie d'Adolf ; l'existence d'Adolf-Pluton n'est donc pas essentielle à la pleine survie d'Adolf. De même,

seul Adolf-Terre aurait pu exister, et cela n'aurait aucunement compromis la survie d'Adolf ; l'existence d'Adolf-Pluton n'est donc pas essentielle à la pleine survie d'Adolf. Puisque ni l'existence d'Adolf-Pluton ni celle d'Adolf-Terre n'est essentielle à la pleine survie d'Adolf, il s'ensuit logiquement que l'existence d'Adolf-Pluton et d'Adolf-Terre n'est pas essentielle à la pleine survie d'Adolf. Autrement dit, Adolf-Pluton ainsi qu'Adolf-Terre auraient très bien pu ne pas exister, sans que cela compromette la survie d'Adolf. Le seul moyen de résoudre ce paradoxe consiste à affirmer que la disjonction de deux faits accidentels puisse être essentielle. Ce n'est pas irrationnel. C'est même le principe de fonctionnement des circuits électriques OU. Le problème est ailleurs, dans la définition de la survie. La définition de Parfit heurte le sens commun, et il n'y a pas de justifications à cela, si ce n'est de permettre à son système de tenir. On s'accorderait mal, en effet, à considérer qu'un organisme détruit à 99% a survécu, même s'il reste 1% de tissus vivants. Or, supposons que l'on copie une personne A en 100 individus différents puis que l'on en tue 99, A aurait pleinement survécu, quand bien même 99% de ce qui le constitue aurait été détruit. Un parfitien pourrait soutenir que A n'a pas pleinement survécu, qu'il est quasiment mort, mais cela irait contre le sens commun puisque au départ comme à la fin du processus, on a un seul individu, qualitativement identique, et cet individu se sentirait bien vivant en chaque point du processus. On pourrait considérer aussi qu'après duplication en 100 individus, A a survécu à 10 000%. Donc, après le retour à un seul individu, on retombe sur un 100% satisfaisant. Ce qui l'est moins, c'est une survie à 10 000%. Qu'est-ce que cela peut bien vouloir dire ? Autant on peut comprendre ce qu'est une survie de moitié ou de quart, par rapport à une moyenne qui correspondrait à, disons, une santé normale, autant il semble obscur de parler de survie démultipliée par 100. Certes, on peut dire : « Je me sens plus vivant que jamais », mais de là à se sentir 100 fois plus vivant que jamais, il y a une

marge considérable… Mais de toute façon, rien de particulier ne serait senti, ce qui souligne une nouvelle fois l'inadéquation de cette conception.

**(b) La survie comme question de degré : l'expérience du spectre combiné :**

Un autre problème est relatif au degré de connexité psychologique qui doit exister entre un individu A existant en t1 et un individu B existant en t2 pour que ces deux individus soit la même personne, ou pour que la personne existant en t1 à travers l'individu A puisse survivre en t2 à travers l'individu B. Parfit contourne le problème de définition d'un seuil, forcément arbitraire, et impliquant un passage très subtil entre des destins complètement différents (vie ou mort), en affirmant que la survie, en plus d'autoriser des relations « *one-to-many* » est une question de degré, pas de tout ou rien. Cette conception est illustrée par des expériences de pensée où une personne – par exemple lui-même, Derek Parfit – est modifiée très progressivement, psychologiquement et/ou physiquement, jusqu'à ressembler à une autre personne – par exemple Greta Garbo[125]. Ces expériences de pensée, Derek Parfit les appelle « expériences du spectre »[126]. Il imagine diverses situations où le processus est mené à 0%, à 10%, à 20%, etc. jusqu'à 100%. Puis il montre que si à un bout du spectre, ainsi qu'à l'autre bout, l'individu résultant est clairement identifiable, respectivement comme Derek Parfit ou comme Greta Garbo, les degrés intermédiaires posent problèmes. On pourrait être tenté de dire : jusqu'à 50% de transformation,

---

125 Actrice suédoise (1905-1990) ayant jouée, entre autres, dans *La Femme aux deux visages* (1941), de George Cukor.

126 *Spectrum experiment*. Il la décline en plusieurs variantes dont : le *physical spectrum*, le *psychological spectrum* et le *combined spectrum*. C'est cette dernière variante, qui comme son nom l'indique, combine modifications physiques et psychologiques, que nous analysons ici.

l'individu résultant est pleinement Derek Parfit, et au-delà, il est pleinement Greta Garbo. Mais qui est-il lorsque la transformation est très exactement à 50% ? Et comment concevoir que quelques modifications très subtiles dans un sens ou dans un autre pourraient avoir une répercussion si considérable : tuer Derek Parfit et faire renaître Greta Garbo – ou l'inverse si on imagine un processus dans le sens contraire ? Ce qui choque ici Derek Parfit semble pourtant commun dans la nature, où des phénomènes chaotiques émergent : à partir de conditions initiales presque identiques – mais néanmoins légèrement différentes – on aboutit à des situations très divergentes au final. Mais il faut garder à l'esprit que justement il est vrai que, par hypothèse, l'individu à 49,9% de transformation est très ressemblant (physiquement est psychologiquement) à l'individu à 50,1% de transformation. Les deux se ressemblent bien plus entre eux que le premier ne ressemble à Derek Parfit ou que le second ne ressemble à Greta Garbo. La différence objective est très faible. La seule supposée différence significative n'est censée affecter que l'identité personnelle. La source d'étonnement réside en ceci, que ce qui est censé changer radicalement n'est pas du domaine de l'observable. C'est comme si, nous classions des verres d'eau selon que leur température est inférieure ou supérieure à 41,2°C. Cela n'a pas de sens, car aucun changement d'état n'a lieu à cette température (le point de congélation est à 0°C et celui d'ébullition à 100°C). Du moins aucun changement apparent, car on pourrait très bien imaginer un organisme vivant aquatique qui pourrait vivre dans une eau de température inférieure à 41,2°C et mourrait dans une eau de température supérieure. Ce serait donc que l'eau, pour cet organisme vivant, aurait une qualité très différente selon qu'elle dépasse ou non les 41,2°C. Prenons le problème différemment. Derek Parfit pense que la survie est une question de degré car dans son expérience du spectre, une différence objective triviale serait censée engendrer une différence subjective non triviale, ce qu'il ne conçoit pas. Mais reprenons notre expérience de pensée de duplication. Adolf-Pluton et

Adolf-Terre, si on les mettait l'un à côté de l'autre, juste après leur « création », seraient indiscernables. Certes, ils n'occupent pas le même espace au même moment, mais qu'on les laisse seuls dans une pièce quelques minutes, et que l'on repasse les voir, on ne saura plus lequel est lequel car entre temps ils ont très bien pu échanger leur place. Objectivement, on peut dire que la différence entre eux deux est très faible. Et pourtant, le sens commun – le même sens commun censé être mis à mal par l'expérience du spectre – n'aura aucun mal à les considérer comme deux personnes différentes, et non pas comme la même personne en deux exemplaires ou comme une seule personne en deux individus.

**(c) Survie branchée, survie en degré :**

A ce propos, peut-on dissocier la conception de la survie comme possiblement « *one-to-many* » de la conception de la survie comme question de degré ? Il semble que non. Imaginons que nous considérions la survie comme une question de degré mais que nous pensions que c'est une relation nécessairement « *one-to-one* ». Alors nous pourrions imaginer une expérience du spectre impliquant Adolf-Pluton et Adolf-Terre. Mais que l'on parte de l'un ou de l'autre, à aucun moment, dans le processus nous ne constaterions un changement objectif, puisque les deux individus sont qualitativement identiques. Donc, admettons que nous partions d'Adolf-Pluton, à 10% de transformation, nous serions toujours fondé à dire : « il s'agit toujours à 100% d'Adolf-Pluton ». Idem à 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% et 100%. Imaginons à présent que nous considérions que la survie est une question de tout ou rien mais qu'elle peut être une relation « *one-to-many* ». Nous pouvons imaginer, comme tout à l'heure, une situation où une personne se trouve répliquée en 100 exemplaires. A ce stade, elle est donc constituée de 100 individus indépendants. Puis, 99 d'entre eux sont tués. Puisque 99% de ce qui constitue notre

personne est alors détruit et si la survie est une question de tout ou rien, alors nous pouvons considérer que notre personne de départ est morte. Pourtant, il semble qu'elle vive toujours ! C'est donc avec une certaine logique interne que Derek Parfit soutient la double thèse suivante : la survie est une question de degré, la survie peut être une relation « *one-to-many* ». Mais cette double thèse a des conséquences désastreuses pour le sens commun. Qui plus est, les arguments qui la soutiennent sont contestables.

**(d) Une interprétation probabiliste :**

Admettons cependant que l'argument du spectre soit valide, c'est-à-dire qu'il n'y a pas de frontière valable, quelque part dans le processus de transformation de Derek Parfit en Greta Garbo, en deçà de laquelle Derek Parfit existe pleinement et Greta Garbo n'existe pas du tout, et au-delà de laquelle Greta Garbo existe pleinement et Derek Parfit n'est plus du tout. Il reste que la solution proposée par Parfit : la survie est question de degré, n'est pas la seule envisageable. On peut aussi imaginer qu'à 0% de transformation en Greta Garbo, l'individu résultant a 100% de chances d'être Derek Parfit, qu'à 10%, il a 90% de chances d'être Derek Parfit, etc. et qu'à 100% de transformation, il a 100% de chances d'être Greta Garbo. On peut donc proposer une interprétation probabiliste de l'argument du spectre, qui n'est pas sans rappeler l'interprétation probabiliste que propose Bruno Marchal des expériences de pensée de duplications à l'identique.

***(2) Considérer, en dépit des apparences, que B et C sont alternativement dans le temps, la même personne (A) puis une autre personne. B et C sont donc 50% A chacune, mais en moyenne sur une***

***certaine durée de temps, et non pas à chaque instant.***

Nous passerons très rapidement sur cette option car, à notre connaissance, elle n'a été défendue par aucun philosophe majeur. Nous la mentionnons seulement parce qu'elle est une option théoriquement possible pour conceptualiser les expériences de pensée de duplication. Elle est une sorte d'hybride entre l'option précédente et l'option 2.a dont elle cherche à concilier les qualités. Mais elle hérite aussi de leurs défauts. Qui plus est, elle pose des questions qu'elle ne résout pas : durant combien de temps d'affilée B ou C sont-ils A ? Comment et pourquoi se passe le changement ? Il semble qu'il faille faire intervenir un « *further fact* » telle qu'une sorte d'ego cartésien pour donner du sens à cette option bizarre.

***(3) Considérer la thèse de l'occupation multiple : L'individu A n'est pas une seule personne, mais deux (B et C) qui finissent par mener une existence séparée.***

Voilà une autre option qui peut paraître bizarre à première vue[127]. Elle est néanmoins défendue par quelques rares philosophes de premier plan. Examinons tout d'abord les problèmes qu'elle pose. Déjà, si l'individu A avait été répliqué trois fois, il semblerait alors, suivant cette hypothèse, qu'il aurait été trois personnes et non pas deux. S'il avait été répliqué *n* fois, il serait *n* personnes et non pas trois ou deux, etc. Si ce que nous croyons être une personne, mettons Paul, est assis dans un fauteuil, comment savoir le nombre réel

[127] Cf. PERRY, John. *Can the Self Divide?* Et : LEWIS, David Kellogg. "Survival and Identity". *In:* RORTY, Amélie Oksenberg. *The Identities of Persons*

de personnes qui sont assises dans ce fauteuil ? Il semble que ce nombre dépende de ce qui arrivera plus tard. C'est-à-dire que si Paul mène une vie ordinaire jusqu'à sa mort, alors il n'y a qu'une seule personne dans le fauteuil. S'il subit une duplication un jour ou l'autre, alors il y a deux personnes assises. S'il subit une réplication donnant lieu à la reconstitution de *n* exemplaires de lui-même, alors il y a *n* personnes assises dans ce fauteuil. Mais comment croire qu'un événement futur hypothétique puisse avoir une influence sur des faits passés ? Le second problème consiste à donner un sens à la coexistence de deux ou plusieurs personnes en ce qui n'apparaît être qu'une seule personne.

**(a) Occupation multiple et perdurantisme :**

Ces deux problèmes peuvent être réglés ensemble, en s'appuyant sur une conception perdurantiste de la personne, l'idée selon laquelle les personnes sont constituées de parties temporelles (John le 23 mars 2006, John aujourd'hui, John en 2008, etc.), de la même manière dont elles sont constituées de parties spatiales (jambes, mains, tête, etc.). Si, lorsque l'on compte le nombre de personnes assises dans un fauteuil, on compte le nombre de *stages temporels de personne*, alors on a raison de dire que seulement une personne est assise dans le fauteuil. Si, lorsque l'on compte le nombre de personnes assises dans un fauteuil, on compte le nombre de *vers 4D*[128] constitués par une succession de stages temporels de personne, alors ce nombre dépend de ce qui va se passer plus tard pour cette personne ou,

[128] Si on projette l'espace-temps, formé de trois dimensions d'espace, et d'une dimension de temps, sur un espace en trois dimensions, où une dimension matérialise la dimension temps, tandis que les deux autres dimensions sont une projection bidimensionnelle des trois dimensions d'espace, alors les choses et les personnes, selon une ontologie perdurantiste, en viennent à ressembler à des sortes de vers de terre, formés d'une chaîne de parties ou phases temporelles agglomérées les unes aux autres, d'où l'expression de « ver 4D » ou « ver quadridimensionnel ».

dit autrement : ce nombre dépend de ce qu'il se passe dans une partie future du temps. Cette situation est comparable à celle où deux routes fusionnent momentanément pour n'en former plus qu'une. Selon une certaine manière de parler, qui tient compte de la situation globale des deux routes, on a bien deux routes, quand bien même elles fusionnent. Selon une autre manière de parler, on ne traverse, en marchant sur cette partie de route fusionnée, que sur une seule bande d'asphalte. Une analogie éclairante peut être donnée à travers le cas des frères siamois. Un être humain bicéphale est aussi deux personnes partageant un même corps, comme l'atteste le fait que les deux têtes peuvent développer une personnalité et un caractère différent. Si on ne voit que son thorax et son abdomen, on pourra penser qu'il ne s'agit que d'une seule personne. Similairement, deux personnes existant à un temps t2 pourraient partager une même partie temporelle en t1. Si on ne voit que cette partie temporelle en t1, on pourra penser qu'il ne s'agit que d'une seule personne.

L'expérience de pensée de fusion d'individus permet également de comprendre comment un seul individu peut être deux personnes. Supposons deux personnes qualitativement identiques et numériquement distinctes[129] : A et B, dont les hémisphères cérébraux sont équipotents[130] (cela arrive réellement parfois).

---

[129] Il semble que Parfit ne devrait pas y voir deux personnes mais une personne à travers deux individus. Toutefois, le sens commun concèdera volontiers que si ce sont bien deux personnes qui se ressemblent comme deux gouttes d'eau, cela n'en fait pas pour autant une seule personne existant à travers deux individus.

[130] **Hémisphères équipotents :** L'équipotence des hémisphères est le contraire de la latéralisation. Elle désigne une latéralisation faible ou nulle d'une ou de plusieurs fonctions cérébrales chez quelqu'un. Par exemple, une personne chez qui les aires du langage sont à peu près équitablement réparties entre l'hémisphère droit et l'hémisphère gauche a des hémisphères équipotents. Dans notre expérience de pensée, nous supposons une équipotence extrême touchant toutes les fonctions cérébrales, autrement dit, une latéralisation nulle de l'ensemble des fonctions du cerveau.

En t1 : L'hémisphère droit de A et l'hémisphère gauche de B sont détruits (on peut imaginer que le processus est très progressif) – N.B. : des considérations cliniques tendent à montrer que c'est compatible avec la survie, notamment si les hémisphères sont équipotents.

En t2 : Les hémisphères restants sont réimplantés dans un individu C, qualitativement identique à A et B, mais anencéphale (là aussi, on peut imaginer que le processus est très progressif). Quelle(s) personne(s) est (sont) l'individu C après opération ?

En t3 : On remet les hémisphères à leur place dans les individus A et B d'origine (là encore, on peut imaginer que le processus est très progressif ; et on peut même imaginer que l'on reconstitue les hémisphères manquants à partir des hémisphères restants). Quelle(s) personne(s) est (sont) les individus A et B à la fin ?

Intuitivement, il semble que l'on sera plutôt conduit à considérer que les personnes en t3 sont les mêmes que les personnes en t1. A savoir A et B. Après tout, la situation en t3 est qualitativement identique à la situation en t1 et entre temps, rien de traumatisant n'a été subit. Certes, on a d'abord un hémisphère qui est détruit pour chaque individu, mais cette destruction est supposée très progressive. Par exemple, on pourrait imaginer que toutes les heures, une cellule cérébrale disparaisse qui n'est pas remplacée. Il va relativement de soi que ce n'est pas la disparition d'une seule cellule cérébrale qui va contrevenir gravement à l'identité numérique d'une personne. Naturellement, il peut arriver qu'une cellule cérébrale meurt, et cela passe entièrement inaperçu. De plus, chaque hémisphère est supposé équivalent à son hémisphère complémentaire, c'est-à-dire que l'on postule que toutes les fonctions cognitives, la mémoire, la personnalité sont réparties de manière homogène entre

l'hémisphère droit et l'hémisphère gauche, de sorte que l'absence d'un des deux hémisphère n'altère pas profondément le psychisme. Ensuite, il y a une déconnexion de chaque hémisphère restant de son corps d'origine, ainsi qu'une double reconnexion : cette dernière consiste en une connexion des hémisphères restants entre eux et, en une connexion de ces derniers avec un nouveau corps. Comme ces processus sont encore une fois très progressifs et que les hémisphères sont qualitativement identiques à leur homologue détruit, il n'y a encore pas lieu de suspecter un quelconque changement dans l'identité numérique. A vrai dire, il n'y a toujours pas de changement notable sur un plan qualitatif, si ce n'est des déplacements d'hémisphères, des déconnexions et des reconnexions, toutes opérations qui – par elles-mêmes seulement – n'altèrent en rien la structure interne des hémisphères. En l'absence de changements qualitatifs majeurs, pourquoi supposer un changement numérique, alors même que des changements qualitatifs majeurs chez une personne (croissance, vieillesse, expérience, etc.) sont supposés préserver l'identité numérique ? Puis, finalement, avec toujours la même progressivité, la même patience, sans rien brusquer, on réintègre chaque hémisphère dans son corps d'origine, puis, progressivement, on reconstruit l'hémisphère manquant. On aboutit ainsi à une situation qualitativement très similaire à la situation de départ. Si nous faisons le bilan de tout cela, il s'impose avec la force de l'évidence qu'à aucun moment, d'un bout à l'autre de notre expérience de pensée, l'identité numérique de nos deux personnes A et B ne s'est perdue ou altérée. Et pour cause, l'identité qualitative s'est préservée également en cours de route, à quelques détails près qui ne sauraient constituer un appui en faveur d'une modification de l'identité numérique, à moins d'admettre que n'importe quelle petite perturbation qualitative suffit à la compromettre. Mais si cela devait être le cas, alors le concept même d'identité numérique personnelle n'aurait plus guère de sens, car les changements importants affectant notre conscience, notre humeur, notre mémoire, etc. d'un jour sur

l'autre, feraient que nous ne serions – littéralement – plus la même personne du jour au lendemain.

En quoi cette expérience de pensée que nous venons d'exposer et d'analyser nous éclaire quant à la thèse de l'occupation multiple ?

Elle nous permet tout d'abord de comprendre comment deux personnes numériquement distinctes (dans notre expérience de pensée : A et B) pourraient coexister au sein d'un même individu, quand bien même nous avons l'impression de n'avoir à faire qu'à une personne.

Elle nous permet de comprendre qu'il n'est pas nécessaire que ces deux personnes soient qualitativement différentes pour qu'elles soient numériquement distinctes. Il suffit qu'elles soient susceptibles de pouvoir exister indépendamment l'une de l'autre, de mener des carrières, des trajectoires de vie différentes, durant des périodes de temps plus ou moins longues, fut-ce de manière intermittente.[131]

---

131 A vrai dire, bien que la thèse de l'occupation multiple et le perdurantisme semblent, de prime abord contre-intuitifs, on peut en proposer une interprétation mathématique, en terme de fonctions, qui les éclairent soudainement d'une aura de plausibilité et les rendent très satisfaisants pour l'esprit. Il suffit que deux fonctions, $f$ et $g$, diffèrent en un seul point, en une seule valeur de $x$, pour que ces deux fonctions soient numériquement différentes, pour tout $x$, même si elles prennent des valeurs semblables pour toutes les valeurs de $x$ sauf une. Par exemple, la fonction $f$, telle que $f(x) = 1$, est une fonction valant 1 pour tout $x$ réel. La fonction $g$, telle que $g(x) = x/x$, est une fonction valant 1 pour tout $x$ réel **fini et non nul**. Ces valeurs de $x$ singulières où $g$ et $f$ diffèrent suffisent à démontrer : $f \neq g$. De même, il suffit que deux personnes soient différentes en un seul moment de leur existence (qualitativement, spatialement, matériellement, etc.) pour qu'elles soient numériquement distinctes en tout point du temps. Souscrire à la fois au perdurantisme et à la thèse de l'occupation multiple c'est, finalement, conceptualiser l'identité d'une personne à la manière d'une fonction mathématique.

**(b) Ontologie du perdurantisme :**

Que doit-on appeler authentiquement *une personne* ? Un stage temporel ou l'ensemble des stages temporels successifs formant ce que les philosophes perdurantistes appellent un « ver temporel » ? Certains auteurs sont plutôt partisans d'une ontologie tridimensionnelle et mettront donc l'accent sur des stages temporels d'extension temporelle nulle, qui sont donc des entités tridimensionnelles non persistantes, c'est-à-dire des événements. D'autres auteurs sont partisans d'une ontologie quadridimensionnelle et mettront la primauté ontologique sur le « ver temporel ». On peut vouloir préférer une ontologie 3D, plus intuitive à certains égards, notamment parce qu'elle échappe au problème de « surpopulation ». Mais elle bute contre une autre intuition puisque les éléments de cette ontologie sont des événements, ce que l'on accepte difficilement pour des personnes.

**(c) Occupation multiple et libre-arbitre :**

Peut-on adhérer en même temps à la thèse de l'occupation multiple et croire au libre-arbitre ? Il semble, de prime abord, y avoir deux difficultés à concilier ces vues. Premièrement, le libre-arbitre semble être une faculté, si elle existe, qui soit strictement personnelle. On ne peut pas être libre à ma place. Or, dans la perspective de l'occupation multiple, qui serait libre ? Soit il faut admettre qu'une des personnes cohabitante impose sa volonté à l'ensemble des autres qui ne sont donc pas libres. Soit, par le plus grand des hasards, il se trouve qu'elles veulent toutes la même chose en même temps. Soit c'est l'ensemble formé par toutes les personnes cohabitantes qui est doté d'un libre-arbitre propre, mais l'ensemble formé par toutes les personnes cohabitantes n'est aucune de ces personnes en particulier. Toutefois, il est possible que toutes soient libres en même temps,

malgré tout. Certes, toutes ces libertés ne peuvent s'actualiser car elles entrent en conflit, et on peut supposer que des rapports de forces se mettent en place. Phénoménologiquement, certaines de nos prises de décision sont difficiles, il semble se mener un véritable combat interne entre diverses volontés contradictoires. Lorsque l'on agit – par hypothèse librement – on a souvent conscience que l'on aurait pu agir différemment, et parfois on exprime un regret. Il manque peut-être un « si ». En effet, nous avons agi de telle façon et pas d'une autre, donc dire que l'on aurait pu agir différemment, sans autre forme de précision, peut sembler incomplet. Dire que l'on aurait pu agir autrement si nous avions agi autrement est une évidence logique qui n'apporte pas d'information supplémentaire. Nous pourrions dire que nous aurions agi autrement, si le rapport de force entre des volontés libres contradictoires avait été différent. Quoi qu'il en soit, il semblerait que la phénoménologie de l'action libre et de la volonté libre soit tout à fait compatible avec l'hypothèse de l'occupation multiple. Il reste quand même un deuxième problème, plus épineux. Le libre-arbitre semble impliquer l'indéterminisme. Or, si l'indéterminisme est vrai, alors le futur est ouvert. Donc, à un certain moment du temps, le nombre de « vers temporels » constitutifs d'une personne (au sens ordinaire du terme) donnée est indéterminé, au sens fort, ontologique, de « indéterminé ». C'est-à-dire que la question : « combien de personnes (au sens quadridimensionnaliste du terme) sont assises sur cette chaise ? », non seulement n'a pas de réponse qui nous soit accessible, mais n'a pas de réponse du tout. Puisqu'il n'est pas encore fixé qu'il y aura, à l'avenir, réplication ou non de la personne considérée, et si réplication il y a, le nombre de « clones » qui seront générés, il n'y a pas encore de réponse fixée à cette question. Ce n'est qu'au moment de la mort que le passé de la personne morte serait définitivement fixé par un curieux phénomène de rétrocausalité. Encore que l'on ne pourrait exclure une possible résurrection en plusieurs exemplaires ! Un partisan cohérent de la thèse de l'occupation multiple semble donc se devoir d'être également partisan du

déterminisme le plus strict, et donc nier le libre-arbitre. Une solution pourrait être la suivante : le nombre de « vers temporels » coexistant au sein d'une personne à un moment donné, n'est pas égal au nombre de duplications qu'elle subira dans le futur, mais au nombre de duplications qu'elle pourrait possiblement subir, au maximum. Malheureusement, ce nombre semble virtuellement infini. Du moins tant qu'on en reste à un critère psychologique de l'identité personnelle. Mais peut-être est-il temps de nous intéresser au critère corporel…

### *B. Critère somatique*

Notre analyse critique du critère somatique sera plus courte que celle que nous avons développée concernant le critère psychologique. Trois raisons à cela : Tout d'abord, le critère somatique a suscité beaucoup moins d'intérêt, au sein de la philosophie contemporaine que le critère psychologique, même si un certain regain semble poindre, notamment à travers des versions radicalisées comme l'animalisme. Ensuite, un certain nombre des problèmes soulevés par le critère somatique ne lui sont pas spécifiques, même si parfois ils ne le frappent pas aussi fortement qu'ils ne frappent le critère psychologique. Enfin, outre ces problèmes communs avec le critère psychologique, le critère somatique présente relativement peu de difficultés supplémentaires. Parfois, ce qui peut apparaître comme une difficulté est plutôt, semble-t-il, un point positif, mais nous y reviendrons. Tout de suite, voyons quels sont les problèmes rencontrés par le critère somatique et auxquels le critère psychologique s'était déjà mesuré. Ces problèmes sont au nombre de deux : le problème de la transitivité, et, dans une moindre mesure, celui de la duplication.

La transitivité pose problème au critère somatique pour une raison comparable au critère psychologique. Qu'est-ce que cela veut dire « avoir le même corps que » ? Le physique, la morphologie, la physiologie, et même la matière constituant le corps, évoluent, à l'instar du psychisme. Le corps pris à un instant t1 n'est plus tout à fait le même corps (qualitativement, et même numériquement, à cause des transferts de matière) qu'à l'instant t2. Fonder une identité numérique de la personne sur une à identité corporelle approximative expose donc à un paradoxe similaire à celui exposé par Thomas Reid à l'encontre du critère de Locke. Mais heureusement, la solution que l'on peut apporter est fort semblable à celle que Parfit propose en réponse à la critique reidienne du critère lockéen. On peut définir une connexité et une continuité somatiques sur le modèle de la connexité et de la continuité psychologiques.

La duplication, évidemment, pose moins de problèmes avec le critère somatique qu'avec le critère psychologique. Pas question d'imaginer qu'il soit possible de téléporter qui que ce soit. Au mieux, l'individu qui apparaîtrait serait une autre personne que celle d'origine. Néanmoins, une variante est possible, plus limitée dans ses extravagances, sous la forme d'expériences de pensée de double transplantation d'hémisphères cérébraux (cf. David Wiggins). Ce genre de scénario de science-fiction a d'ailleurs été largement étudié – avec délice – par Derek Parfit[132]. Il faut dire que le cerveau est à la fois province du corporel et province du psychologique. Traiter son cas permet en quelque sorte de faire d'une pierre deux coups. Aux mêmes maux les mêmes remèdes, ou presque. L'essentiel des réflexions que nous avons tenues au sujet du problème de la duplication dans le cadre du critère psychologique est transposable, modulo quelques changements mineurs parfois, au critère somatique. Pour mémoire, voici les solutions envisageables. Imaginons donc qu'une

132 PARFIT, Derek, *Reasons and Persons*.

personne A dont le cerveau est équipotent voit son hémisphère gauche implanté dans un individu B décérébré et son hémisphère droit implanté dans un individu C décérébré (avec toutes les précautions utiles pour éviter tout traumatisme physique et psychologique…). Question : Où se trouve A ? Pour y répondre, on peut :

1. Assumer, au moins partiellement, la logique de l'identité personnelle, c'est-à-dire admettre que A = B et que A = C. On peut le faire d'au moins deux façons :

    1.1. Considérer que A = B et que A = C et accepter la conclusion logique selon laquelle B = C : C'est-à-dire qu'il faudrait admettre, dans notre ontologie, la possibilité pour deux êtres n'occupant pas ou pas exactement le même espace au même moment d'être numériquement identiques.

    1.2. Considérer que A = B et que A = C et refuser la conséquence logique suivant laquelle B = C : C'est-à-dire qu'il faudrait remettre en question la transitivité de l'identité.

2. Ne pas assumer la logique de l'identité personnelle et suggérer que ce cas constitue une exception. On peut le faire au moins de deux façons :

    2.1. Considérer que A = B ou bien A = C : C'est-à-dire que des deux personnes B et C, au moins une et au plus une, sera la personne A de départ. L'autre est un « imposteur » qui s'ignore. On peut soutenir un tel point de vue d'une manière sensiblement différente d'avec le critère

psychologique, en arguant que contrairement aux apparences imposées par notre expérience de pensée, une équipotence parfaite des hémisphères n'existe pas et que l'un des deux sera dominant et véritable porteur de l'identité personnelle. Dans le cas général où les hémisphères ne sont pas équipotents, on pourrait défendre l'idée selon laquelle c'est l'hémisphère siège de la parole (la plupart du temps le gauche), ou celui avec lequel la personne s'identifie le plus (par exemple le gauche pour un scientifique, le droit pour un littéraire), qui est le siège véritable de l'identité numérique de la personne.

2.2. Considérer que A ≠ B et A ≠ C : C'est-à-dire que B et C sont des « imposteurs » qui s'ignorent. Position que Derek Parfit caricature par « Comment un double succès [chirurgical] pourrait être un échec ? », puisque, en effet, ce genre d'expérience de pensée est couramment précédé du constat de réussite de la greffe de cerveau si elle ne concerne qu'un des deux hémisphères et que l'autre est détruit.

3. Ne pas assumer la logique de l'identité personnelle et suggérer une reconceptualisation « exotique » de la notion de personne :

3.1. Considérer, en dépit des apparences, que B et C sont une même personne (A) existant à travers deux individus. Cette possibilité doit être distinguée de la première. Il ne s'agit pas de dire que B est pleinement A et que C est pleinement A, mais que A est, après la duplication, un être formé

de deux parties spatiales disjointes : B et C, qui sont donc à 50% A chacune.

3.2. Considérer, en dépit des apparences, que B et C sont alternativement dans le temps, la même personne (A) puis une autre personne. B et C sont donc 50% A chacune, mais en moyenne sur une certaine durée de temps, et non pas à chaque instant. Cette position semble cependant encore plus incongrue dans une optique somatique que psychologique…

3.3. Considérer la thèse de l'occupation multiple[133] : L'individu A, en dépit des apparences, n'est pas une seule personne, mais deux (B et C) qui finissent par mener une existence séparée. Par contre, cette dernière thèse jouit d'une soudaine « salvation » dans le cadre du critère somatique, puisque l'on n'est plus obligé d'admettre une « surpopulation ». Le nombre de duplications possibles de chaque personne étant limité grosso modo à deux, le nombre d'hémisphères. On peut donc admettre qu'il y a toujours deux personnes et seulement deux en ce que nous appelons nous-mêmes. Une, liée à l'hémisphère gauche. L'autre, liée à l'hémisphère droit. Du coup, nous ne sommes plus obligés de souscrire à un déterminisme strict et nous pouvons « sauver » le libre-arbitre. Car, rappelons que nous avions été contraints, précédemment, dans notre examen de la thèse de l'occupation multiple dans une perspective psychologique, de rejeter le libre-arbitre et l'indéterminisme. Ceci afin que le nombre de personnes coexistantes en

[133] Cf. PERRY, John. *Can the Self Divide?* Et : LEWIS, David Kellogg. "Survival and Identity". *In:* RORTY, Amélie Oksenberg. *The Identities of Persons*

> chaque individu à un moment donné puisse avoir une valeur définie. Dans la perspective somatique, cette valeur est définie de fait par les contraintes de la corporalité. On ne peut diviser indéfiniment le cerveau d'un individu pour le démultiplier à l'infini. A un moment où à un autre, on tombe sur un niveau minimum de complexité requis pour que l'on puisse encore parler d'un rudiment de conscience ou même simplement de sentiment.

Signalons pour finir, l'existence d'une curieuse convergence qui s'esquisse entre la thèse de l'occupation multiple et l'option 2.1. lorsque celles-ci sont replacées dans le cadre du critère somatique.

### 1. Le critère somatique est-il intuitif ?

Une réelle difficulté posée par le critère corporel est ailleurs. Ce critère semble ne pas saisir ce qui intuitivement paraît être le plus important dans la définition d'une personne : la conscience de soi. Par exemple : un embryon ou un malade d'Alzheimer en phase terminale seraient des personnes[134]. Mais est-ce vraiment le plus important ? Revenons à une expérience de pensée impliquant une téléportation. Jusque maintenant, nous sommes partis du principe que le corps de l'individu d'origine qui entre dans le téléporteur pour aller d'un point à un autre, était *détruit*. Or, ce présupposé a toute son importance pour la crédibilité d'un critère psychologique. Imaginons : suite à un traumatisme, vous perdez quelque peu la mémoire et changez quelque peu de personnalité : par exemple, vous devenez très lent dans

[134] Bien sûr, un individu atteint de la maladie d'Alzheimer en phase terminale sera sans doute *considéré comme* une personne par ses proches, pour des motifs d'attachement et d'affection envers la personne qu'elle *a été*. Mais acceptera-t-on facilement de dire que l'essentiel d'une personne réside en un organisme biologique végétant ?

vos mouvements et vos réactions, et quelque peu d'humeur aigrie. Au même moment, quelque part, sur une planète tournant autour de Proxima du Centaure, une civilisation avancée créé un individu en tout point indiscernable de la personne que vous étiez avant votre accident. Ce nouvel individu jouit donc d'une bien meilleure continuité psychologique avec la personne que vous étiez, que la personne que vous – en tant qu'être humain – êtes devenue. Pourtant, il semble hautement contre-intuitif d'imaginer que vous – d'avant – n'êtes plus cette personne diminuée – que tous vos proches aimeront et identifieront bien comme étant vous-même – mais que vous vous retrouvez soudainement projeté à 43 000 milliards de kilomètres de la Terre… Ici, il semble clairement que le critère corporel prime sur le critère psychologique.

A travers cette expérience de pensée transparaît un travers gênant du critère psychologique : il nous interdit de changer ou du moins restreint la latitude de notre changement, il rend incompréhensible ou moins évident le concept d'oubli, notamment lorsqu'il concerne les croyances égologiques. Donc, en dépit de l'attrait premier qu'il peut susciter, il est finalement moins en accord avec le sens commun que le critère somatique. Considérons encore ceci : Un homme tue délibérément un autre homme, gratuitement, par pure méchanceté. Suffit-il qu'il se distancie de l'état d'esprit qui l'animait alors, qu'il ne comprenne plus pourquoi il a agi ainsi et condamne son crime pour que, d'un coup de baguette magique, il soit absout de ce crime, dans la conscience des gens et aux yeux de la loi ? Et pire, cela suffit-il pour ôter de lui toute responsabilité morale ?

## 2. L'amnésie antérograde

Une deuxième difficulté cible plus particulièrement le concept de *continuité psychologique*. Nous l'avons vu, Derek Parfit a défini le concept de connexité et de continuité psychologique, afin de répondre à l'objection de Thomas Reid contre le critère de l'identité personnelle de John Locke. Mais une autre objection, inspirée notamment par certains cas cliniques mais aussi des considérations sur le sommeil, n'est malheureusement pas sauvée par ces notions parfitiennes.

Certains cas cliniques, comme celui du patient H.M. étudié par Scoville et Milner[135], sont pour le moins curieux. H.M. a subi une opération du cerveau pour tenter de limiter les conséquences de l'épilepsie dont il souffrait. Après cette opération, H.M est devenu sévèrement amnésique. Il se souvenait encore d'événements de sa vie passée – de son enfance jusqu'à quelques heures avant l'opération – mais ne pouvait plus retenir d'informations nouvelles. Un psychologue pouvait par exemple venir le matin lui faire subir une batterie de tests, il ne le reconnaîtrait plus l'après-midi et n'aurait aucun souvenir de la matinée. En revanche, H.M. présentait des performances tout à fait normales aux tâches d'empan mnésique ce qui suggère une mémoire à court terme intacte.

Supposons une personne normale allant se coucher. Lorsque le lendemain matin elle se réveille, elle peut se remémorer des éléments importants de sa journée de la veille, et organiser sa journée de manière à accomplir certains objectifs qu'elle s'est fixée la veille. C'est un minimum qui la plupart du temps s'accompagne de souvenirs bien plus anciens.

---

135 SCOVILLE, William Beecher and MILNER, Brenda. "Loss of recent memory after bilateral hippocampal lesions".

Lorsque l'on se réveille, en général, au bout de quelques minutes, on est en mesure de se remémorer les étapes les plus importantes de sa vie. Qui sommes-nous, d'où venons-nous, quels sont nos parents et amis, que voulons-nous, quels sont nos buts, etc. ?

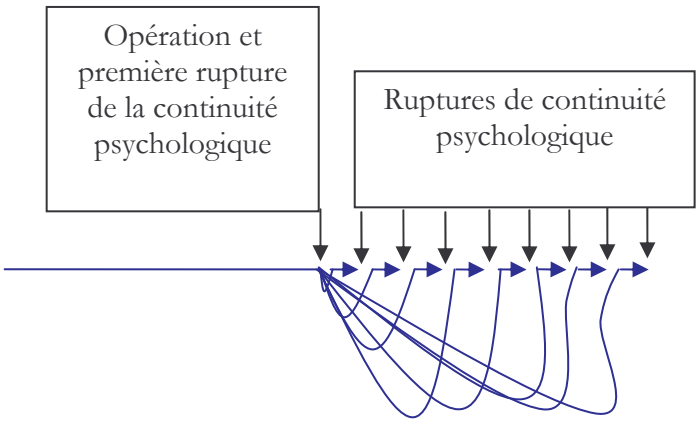
Imaginons toutefois que nous venons tout juste de nous réveiller, et que nous sommes, comme on le dit souvent « dans le gaz », sans souvenir bien précis de notre passé. Techniquement, on ne peut parler de connexité psychologique à ce moment là, puisque nous n'avons aucun souvenir du passé, quand bien même nous conservons la disposition à le remémorer. Et puisque l'on ne peut parler de connexité psychologique, on ne peut parler non plus de continuité psychologique. Doit-on en conclure que nous sommes morts et que nous revenons à l'existence quand, quelques secondes plus tard, nous reprenons conscience claire et distincte de nous-même ?

Mais que dire alors du patient H.M. ? Quand le matin il se réveille, non seulement il ne se souvient absolument pas de la journée de la veille et ne peut s'en souvenir, mais il n'a aucune mémoire non plus de toutes les autres journées qui ont suivi son opération. En suivant la logique de Parfit, on ne peut donc clairement pas parler de connexité psychologique, et a fortiori de continuité psychologique entre une journée de ce patient et la journée suivante. Chaque journée, chaque heure, et même chaque minute est toujours, pour lui, la première, après un sommeil sans rêve d'une durée indéterminée ayant suivie son opération. On peut dire que, d'une certaine façon, le patient H.M. vit dans un éternel présent, mais le présent de l'instant, où aucun souvenir postérieur à l'opération et aucun projet de vie pour l'avenir n'est possible.

Toutefois, comme nous l'avons dit, H.M. se souvient encore d'événements de sa vie d'avant l'opération. En toute rigueur, nous devrions donc affirmer qu'à tout moment, H.M. est en connexité psychologique avec ce passé, et donc en continuité psychologique avec lui.

Nous nous retrouvons ainsi dans une situation paradoxale : Depuis son opération, aucune journée de la vie de H.M. n'est en continuité psychologique avec une autre. Depuis son opération, toutes les journées de la vie de H.M. sont en continuité psychologique avec sa vie passée.

Si l'on représente cela schématiquement, nous obtenons quelque chose comme ceci :




Continuité psychologique du patient H.M.

Parfit nous dit que la survie est une question de degré, en relation directe avec la continuité psychologique qui est une question de degré. Toutefois, dans le cas de H.M., il y a à la fois pleine continuité psychologique et totale absence de continuité psychologique. H.M. a donc à la fois pleinement survécu à son opération et entièrement cessé d'exister

suite à son opération. Imaginons que vous ayez connu H.M. avant son opération, puis l'ayez rencontré à nouveau, hier. En discutant avec lui, et en l'entendant parler de sa vie passée, vous n'avez eu aucun doute, H.M. a bien survécu à son opération, c'est bien la même personne que vous avez eu devant vous. Supposez maintenant que vous le revoyez le lendemain. Il n'a plus aucun souvenir de la journée de hier, il y a donc eu rupture de la continuité psychologique entre temps. Toute rupture de la continuité psychologique impliquant une rupture de la connexité psychologique, et la connexité psychologique étant le critère parfitien de la survie, H.M. est donc mort, il a cessé d'exister, entre hier et aujourd'hui. Mais le voilà pourtant qui se remet à vous parler de sa vie passée. Il y a donc connexité psychologique entre cette vie passée et aujourd'hui : H.M. est donc bien vivant. H.M. est donc 100% mort, puisque entre hier et aujourd'hui, il n'y a pas connexité psychologique et pourtant il est vivant, puisque entre sa vie passée et hier et entre sa vie passée et aujourd'hui, il y a une connexité psychologique forte. Comment cela peut-il se faire ? Comment H.M. peut-il être à la fois pleinement mort et pleinement vivant ?

Parfit dit aussi que la survie peut être une relation « *one-to-many* ». On pourrait donc se dire que H.M. d'avant l'opération a survécu à travers plusieurs individus HM1, HM2, HM3, etc. chacun correspondant à, mettons, une journée de la vie de l'être humain H.M. Mais il y a aussi un problème, logique cette fois-ci.

HM1, HM2, HM3, etc. sont-ils vraiment des individus ? Quel est leur statut ontologique ? Selon l'hypothèse de Parfit, ce ne sont pas des personnes, puisqu'ils sont constitutifs d'une personne : H.M. Sont-ce des individus ? Par analogie avec les cas de duplication, on pourrait dire que oui. Mais d'un autre côté, il y a une unité corporelle de l'être humain H.M. depuis sa naissance. Il semble donc qu'il n'y ait qu'un seul individu,

lequel coïncide avec une seule personne. Revenons au cas de Dr Jekyll et Mr Hyde. Selon un critère psychologique, ce sont deux personnes différentes. Maintenant, supposons que tout ne les oppose pas et qu'ils aient une pensée égologique commune. Pour fixer les choses : supposons qu'ils aiment tous deux les salsifis. Cela suffirait-il pour en faire une seule et même personne ? Probablement pas. Mais supposons qu'on leur ajoute une nouvelle pensée égologique en commun, puis une autre, puis encore une autre, etc. Lorsque toutes ou la plupart de leurs pensées égologiques seront communes, on aura une seule et même personne. Mais où se trouve la limite précise qui va faire passer de deux personnes à une seule ? Quel « stock » de pensées égologiques communes est nécessaire ? On se retrouve dans une sorte d'expérience de pensée du spectre, chère à Parfit, et pour une discussion approfondie de ces expériences de pensée, nous vous renvoyons à la partie qui en traite spécifiquement. On pourrait bien sûr suggérer qu'il existe bien une limite tranchée, ou bien suggérer une interprétation probabiliste : plus Jekyll et Hyde ont de pensées égologiques en commun, plus ils ont de chance de n'être qu'une seule et même personne. Mais on sort alors de l'optique parfitienne ! Si on veut y rester, on est obligé de dire que la question de savoir si Jekyll et Hyde sont une seule et même personne, et donc s'ils sont des personnes tout court, est indécidable lorsque le pourcentage de pensées égologiques partagées par rapport à la quantité totale de pensées égologiques avoisine 50%. Mais dans cette situation, Jekyll et Hyde sont, dans leurs différences, et par rapport à un bloc de pensées égologiques communes que l'on pourrait nommer « Jekyll-Hyde », dans une situation assez comparable à HM1, HM2, HM3, etc. par rapport à H.M. d'avant l'opération. Dans ce dernier cas, on ne peut pas plus trancher, concernant le statut ontologique de HM1, HM2, HM3, etc. que dans le cas de Jekyll et Hyde tel que nous l'avons modifié.

On peut présenter encore cela autrement. Supposons qu'après votre mort, Dieu vous ressuscite en recréant un être qui soit exactement similaire à l'individu que vous étiez au moment de votre trépas. D'après le critère de connexité psychologique, cet individu est vous-même. Mais supposons qu'il ressuscite un être qui soit parfaitement semblable à ce que vous étiez à l'âge de 10 ans (et avec la psychologie et les souvenirs que vous aviez alors), vous a-t-il redonné la vie ? Il semble que non. Vous n'êtes pas cet être de 10 ans que vous avez, certes, été un jour. La raison en est qu'entre-temps vous avez profondément changé. Ce changement, si on en croit le critère psychologique, est constitutif de ce que vous êtes à présent et conditionne ce que vous pourrez être plus tard. Imaginons qu'à 11 ans, vous ayez subi une duplication à l'identique. Vous avez ensuite mené une vie juste et dévouée au Christ. Inversement, votre contrepartie a ensuite méprisé Dieu et s'est adonnée au vice. Vous êtes devenu clairement deux personnes différentes aux yeux de tous. En ressuscitant votre ancienne version commune âgée de 10 ans, qui Dieu a-t-il réellement ressuscité ? Celui (vous) qui mérite le Paradis ou celui (votre contrepartie) qui mérite l'Enfer ou encore quelqu'un d'autre ? Impossible de répondre. Pourtant cet enfant est en connexité psychologique avec l'enfant que vous étiez à 10 ans.

Avec H.M., on est face au même problème. HM1 a modifié H.M. d'origine. HM2, lorsqu'il apparaît est un avatar, non pas du HM modifié par HM1 mais d'une version antérieure, pré-HM1, qu'il remodèle autrement. HM3, lorsqu'il paraît, incarne une troisième fois pré-HM1. Mais qui est-il ? HM1 ? HM2 ? Quelqu'un d'autre ? Là aussi, il est impossible de trancher.

### 3. Critère somatique et présent spécieux

Pour résoudre ce paradoxe, nous pouvons enlever une contrainte au concept de *connexité psychologique* et dire que deux états psychologiques sont connexes, si et seulement si au moins une partie de l'un est aussi une partie de l'autre. Cela implique moins, semble-t-il, que le fait, par exemple, de réussir à se remémorer un q-souvenir. Il s'agit plutôt d'une unité de conscience partielle (ou totale) et diachronique, comme lorsque l'on écoute un morceau de musique, où l'on ne prend pas conscience des notes individuellement et séparément, mais où l'on prend conscience de la mélodie directement, comme un tout vécu au présent. Ce que le patient H.M. est d'ailleurs parfaitement en mesure de faire. Un autre fameux cas d'amnésie antérograde sévère concerne d'ailleurs un musicien, Clive Wearing, victime d'une encéphalite ayant largement détruit son hippocampe ainsi que quelques zones de son cortex cérébral (dans son lobe temporal gauche et une partie de son lobe frontal).

Clive Wearing, comme H.M., présente une totale incapacité à emmagasiner de nouveaux souvenirs. Il a continuellement l'expérience « d'être conscient pour la toute première fois ». Il ne cesse d'ailleurs de coucher par écrit cet événement extraordinaire parce qu'il croit qu'il est très important. A chaque fois qu'il voit sa femme, il fond en larmes d'émotion comme s'il ne l'avait pas vue depuis des années[136]. Comme pour le patient H.M., il n'y a ni passé ni futur pour lui, juste une existence fragmentée en moments. Malgré ce handicap sévère, Clive Wearing demeure capable de jouer du piano remarquablement bien.

---

[136] Cf. le témoignage de cette dernière, dans : WEARING, Deborah. *Forever Today.*

L'idée selon laquelle le présent ne serait qu'un point mathématique sans extension est une vue de l'esprit qui ne correspond ni à la réalité vécue, comme le montre Francis Kaplan[137], ni à la réalité objective.

En philosophie et en psychologie, ce présent étalé se nomme « specious present » ou présent spécieux, apparent, psychologique (chez William James[138], Bertrand Russell[139], Francis Kaplan, etc.).

Il semble correspondre au concept, issu de la psychologie cognitive, de mémoire sensorielle, registre sensoriel ou buffer sensoriel.

Il s'agit de la structure qui garde pendant un très court laps de temps l'information sensorielle (sons, images, odeurs, etc.) qui nous atteint, la plupart du temps sans que nous en soyons conscients. A titre indicatif, la durée du registre visuel (ou iconique), par exemple, est inférieure à 0,5 secondes, celle du registre sonore (ou échoïque) peut atteindre 2 secondes. C'est grâce à cette structure que nous pouvons percevoir le mouvement, la fluidité, etc. et que, d'une manière générale, le présent a son épaisseur. Par exemple, si les notes d'une mélodie sont séparées par une durée inférieure à 2 secondes (durée maximale du registre sonore), elles seront effectivement perçues comme un tout mélodique, tandis que si la durée qui les sépare est supérieure à 2 secondes, elles ne seront pas perçues comme les notes d'une mélodie.

---

137 Cf. KAPLAN, Francis. « L'irréalité de la mort », *In* : LENOIR, Frédéric et DE TONNAC, Jean-Philippe, (Eds.). *La Mort et l'Immortalité. Encyclopédie des Savoirs et des Croyances*

138 Cf. JAMES, William. *The Principles of Psychology*

139 Cf. RUSSELL, Bertrand. "On the Experience of Time"

Cette fenêtre de 2 secondes maximales correspondant à la durée du registre échoïque ne se déplace pas par saut de 2 secondes au cours du temps, mais en continu. C'est pourquoi une mélodie nous apparaît comme un tout et non comme un ensemble de paquets mélodiques de 2 secondes ! Chaque note fait partie de plusieurs fenêtres à la fois, ce qui justifie de parler de connexité psychologique à propos du registre échoïque. La même remarque s'applique également aux autres types de registres sensoriels, tel par exemple le registre iconique.

Il semble que le registre sensoriel soit également une composante de ce qu'Antonio Damasio[140] appelle « Proto-Soi ». Neurobiologiquement, le proto-soi est défini par Damasio comme « une collection reliée et temporairement cohérente de représentations neuronales qui représentent l'état de l'organisme, à de multiples niveaux du cerveau ». C'est sur le « Proto-Soi » que s'enracinerait la conscience.

Selon Damasio, cette dernière se subdivise en deux parties.

La première partie est la conscience noyau (*core consciousness*), qui est le simple sentiment d'avoir des sentiments et qu'il existe un sujet (ce que Damasio appelle le *Soi central*) qui les ressent. La conscience noyau et le soi central, tout comme le proto-soi, existent dans le ici-maintenant. Sur eux reposent la mémoire de travail (qui correspond grosso modo à ce que l'on appelle communément la mémoire à court terme) et l'attention soutenue qui permettront le passage des images mentales vers la mémoire à long terme (mémoire procédurale, mémoire sémantique et mémoire autobiographique). Seule une petite partie des informations présentes dans le registre sensoriel sont mises en valeur par la

[140] DAMASIO, Antonio R. *Le Sentiment même de soi*, p. 201-208.

mémoire de travail et susceptibles d'être retenues sur le long terme. Heureusement, car sinon notre mémoire serait submergée d'informations inutiles. Le reste correspond sans doute à ce que Leibniz[141] appelle les « petites perceptions », c'est-à-dire tous ces éléments de l'expérience qui ne passent pas le seuil de la conscience proprement dite (la conscience noyau de Damasio).

Chez H.M. et Clive Wearing, la conscience noyau et le soi central sont relativement intacts. Cela leur permet, non seulement de ressentir qu'ils existent, mais également de le savoir, même si ce savoir ne peut être emmagasiné sur le long terme. Il semble même que le passage d'information de leur mémoire de travail à leur mémoire procédurale est relativement intact. En effet, on a pu montrer qu'ils sont capables d'acquérir de nouvelles habiletés, c'est-à-dire de nouveaux savoir-faire (bien qu'ils n'ont aucune conscience de tous les exercices qu'ils ont dus faire pour les développer).

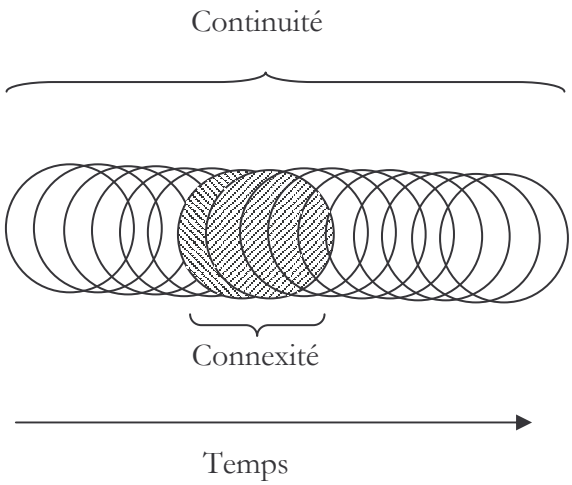
La deuxième partie de la conscience est la conscience étendue (*extended consciousness*) qui se forme sur la base de la mémoire autobiographique. C'est une conscience qui dépasse l'ici-maintenant, vers l'ailleurs, le souvenir du passé et l'anticipation du futur. C'est à ce niveau qu'un projet de vie peut se construire, et que l'on prend conscience de son identité personnelle à travers le temps. Cette identité personnelle à travers le temps doit être comprise comme une connexité psychologique au sens de Parfit. Elle est la quintessence relativement stable et permanente de nos caractéristiques psychologiques, de nos objectifs, de nos idées, de nos souvenirs, bref, de tout ce qui fait de nous le même, socialement et surtout psychologiquement parlant. Tous ces éléments ne sont pas nécessairement présents à la conscience continuellement (et heureusement !) mais ils sont, normalement,

[141] LEIBNIZ, Gottfried Wilhelm, *Les Nouveaux Essais sur l'Entendement Humain*, Préface, p. 41-42.

continuellement disponibles. Ces éléments constituent ce que Damasio appelle le Soi autobiographique.

Dans le cas d'H.M. ou de Clive Wearing, c'est le passage des informations de la mémoire de travail à la mémoire sémantique et autobiographique qui est détruit. C'est donc la conscience étendue et le soi autobiographique qui s'en trouvent profondément altérés, pas la conscience noyau ni le soi central, et encore moins le proto-soi.

Pour plus de clarté, nous pouvons représenter ces notions de manière graphique.




Chaque cercle représente un moment de conscience (au sens de sentiment) unifié *synchroniquement* et *diachroniquement*. C'est le présent spécieux des philosophes, le contenu du registre sensoriel des psychologues, les configurations neuronales temporairement cohérentes du proto-soi de Damasio.

On peut faire l'hypothèse selon laquelle cette *continuité* au niveau du *présent spécieux* est le fondement de l'identité personnelle au sens le plus profond – et numérique – du terme. Dans *Les Nouveaux Essais sur l'Entendement Humain* [142], Leibniz, en son temps, avait déjà développé une conception très proche. A travers le personnage de Théophile, il écrit :

> « Un être immatériel ou un esprit ne peut être dépouillé de toute perception de son existence passée. Il lui reste des impressions de tout ce qui lui est autrefois arrivé et il a même des pressentiments de tout ce qui lui arrivera ; mais ces sentiments sont le plus souvent trop petits pour pouvoir être distingués et pour qu'on s'en aperçoive, quoiqu'ils puissent peut-être se développer un jour. Cette continuation et liaison de perceptions fait le même individu réellement, mais les aperceptions (c'est-à-dire lorsque l'on s'aperçoit des sentiments passés) prouvent encore une identité morale, et font paraître l'identité réelle. La préexistence des âmes ne nous paraît pas par nos perceptions, mais si elle était véritable, elle pourrait se faire connaître un jour. Ainsi il n'est point raisonnable que la restitution du souvenir devienne à jamais impossible, les perceptions insensibles [N.B. : Comprendre les sensations et sentiments inconscients, non perçus, autrement dit ce que Leibniz appelle aussi : « petites perceptions »] (dont j'ai fait voir l'usage en tant d'autres occasions importantes) servant encore ici à en garder les semences. »

Dans la préface des *Nouveaux Essais sur l'Entendement Humain*[143], Leibniz définit la notion centrale de « petites perceptions », qui correspond, semble-t-il, à celle de sensations inconscientes / subconscientes, et qui est donc pour lui la clef de la compréhension de l'identité réelle (sur laquelle repose en second lieu l'identité personnelle) :

---

[142] LEIBNIZ, Gottfried Wilhelm, *Les Nouveaux Essais sur l'Entendement Humain*, Livre II, Chapitre XXVII, § 14, p. 186.

[143] LEIBNIZ, Gottfried Wilhelm, *Les Nouveaux Essais sur l'Entendement Humain*, Préface, p. 41-42.

« Il y a mille marques qui font juger qu'il y a à tout moment une infinité de perceptions en nous, mais sans *aperception*[144] et sans *réflexion*, c'est-à-dire des changements dans l'âme même dont nous ne nous apercevons pas, parce que ces impressions sont ou trop petites et en trop grand nombre, ou trop unies, en sorte qu'elles n'ont rien d'assez distinguant à part; mais jointes à d'autres, elles ne laissent pas de faire leur effet et de se faire sentir au moins confusément dans l'assemblage. C'est ainsi que l'accoutumance fait que nous ne prenons pas garde au mouvement d'un moulin ou à une chute d'eau, quand nous avons habité tout auprès depuis quelque temps. Ce n'est pas que ce mouvement ne frappe toujours nos organes, et qu'il ne se passe encore quelque chose dans l'âme qui y réponde, à cause de l'harmonie de l'âme et du corps, mais ces impressions qui sont dans l'âme et dans le corps, destituées des attraits de la nouveauté, ne sont pas assez fortes pour s'attirer notre attention et notre mémoire, attachées à des objets plus occupants. Car toute attention demande de la mémoire, et souvent, quand nous ne sommes point admonestés[145], pour ainsi dire, et avertis de prendre garde à quelques-unes de nos propres perceptions présentes, nous les laissons passer sans réflexion et même sans être remarquées. Mais si quelqu'un nous en avertit incontinent[146] après et nous fait remarquer, par exemple, quelque bruit qu'on vient d'entendre, nous nous en souvenons et nous nous apercevons d'en avoir eu tantôt quelque sentiment. Ainsi c'étaient des perceptions dont nous ne nous étions pas aperçus incontinent, l'aperception ne venant dans ce cas que de l'avertissement après quelque intervalle, tout petit qu'il soit. Et pour juger encore mieux des petites perceptions que nous ne saurions distinguer dans la foule, j'ai coutume de me servir de l'exemple du mugissement ou du bruit de la mer, dont on est frappé quand on est au rivage. Pour entendre ce bruit comme l'on fait, il faut bien qu'on entende les parties qui composent ce tout, c'est-à-dire les bruits de chaque vague, quoique chacun de ces petits bruits ne se fasse connaître que dans l'assemblage confus de tous les autres ensemble, c'est-à-dire dans ce mugissement même, et qu'il ne se remarquerait pas si cette vague, qui le fait, était seule. Car il faut qu'on soit affecté un peu par le mouvement de cette vague et qu'on ait quelque perception de chacun de ces bruits, quelque petits

144 *Aperception :* fait de s'apercevoir, conscience.

145 Quand nous ne sommes point admonestés : quand on ne nous y fait pas penser.

146 *Incontinent :* tout de suite.

> qu'ils soient; autrement on n'aurait pas celle de cent mille vagues, puisque cent mille riens ne sauraient faire quelque chose. »

L'identité réelle d'un individu est caractérisée par l'enchaînement temporel de ces petites perceptions qui les unit dans une unité substantielle, l'âme. Pour Leibniz, l'identité réelle est une condition nécessaire mais non suffisante à l'identité personnelle qui pour lui n'est qu'une identité de surface, apparente, un symptôme mais non un critère de l'identité réelle. Il s'oppose donc à Locke à deux niveaux. Il réincarne l'âme dans le corps, le monde et autrui en défendant l'idée selon laquelle l'âme possède des contenus non conscients (il se pose ainsi en précurseur de Freud), alors que Locke, en cartésien, poursuit une assimilation de la conscience à l'âme. Contrairement à Locke qui sépare radicalement l'identité personnelle de l'identité de la substance, ce qui lui permet d'imaginer des cas fictionnels où elles ne coïncident pas, Leibniz insiste sur une dépendance de la première à la seconde. Il ne suffit pas, pour fonder l'identité d'une personne, que cette dernière reconnaisse ses souvenirs comme étant les siens, encore faut-il un critère interne qui se trouve dans l'unité d'une substance permise par la connexion réelle entre des perceptions insensibles qui constituent comme la trame indéchirable de son existence. Leibniz va même jusqu'à affirmer que l'identité personnelle, en l'absence de remémoration consciente, peut être sauvegardée par cette continuité ininterrompue de petites perceptions, car elles laissent dans l'âme une trace de leur passage : « en conséquence de ces petites perceptions le présent est gros de l'avenir et chargé du passé ». De plus, le témoignage d'autrui peut venir combler les lacunes dans notre propre souvenir. Ainsi, Leibniz s'oppose à la morale de Locke qui estime qu'on ne doit pas punir une personne pour un crime dont elle n'aurait plus aucun souvenir [147]. C'est d'autant plus compréhensible que Leibniz, s'opposant encore

[147] LEIBNIZ, Gottfried Wilhelm, *Les Nouveaux Essais sur l'Entendement Humain*, Livre II, Chapitre XXVII, § 16, p. 187.

une fois au cartésianisme (cette fois-ci aux concepts de liberté d'indifférence[148] et de liberté absolue[149]) considère que ce que l'on appelle libre-arbitre résulte en fait d'un déterminisme inconscient lié aux petites perceptions.

Outre les arguments et exemples avancés par Leibniz, l'existence des petites perceptions est également une explication possible à des phénomènes comme la cécité aux changements[150], et surtout à la remémoration consciente de faits ayant échappés à la conscience lors de leur perception.

La principale différence entre nos vues et celles de Leibniz est que pour lui, les « petites perceptions » sont toutes stockées dans une sorte de mémoire inconsciente quoique susceptible d'être activée. Comme cela a déjà été souligné à plusieurs reprises par divers auteurs, cela place Leibniz en digne prédécesseur de Freud et de la psychanalyse. Que tout ce que l'on expérimente soit intégralement enregistré quelque part en nous-même quand bien même cela n'a jamais atteint le seuil de notre conscience, est une hypothèse très controversée, essentiellement spiritualiste et sans fondement biologique véritable, même si

---

148 La liberté d'indifférence consiste à prendre une décision lorsque les motifs d'agir d'une façon plutôt que d'une autre sont exactement identiques. On illustre souvent la liberté d'indifférence par son absence dans la fiction de l'âne de Buridan, situé à égal distance d'un seau d'eau et d'un bol d'avoine, et qui, ayant autant faim que soif, se laisse mourir sans aller en premier vers l'eau ou vers l'avoine. Leibniz considère qu'à cause des petites perceptions, il est impossible d'obtenir une situation de conflit de la sorte qui permettrait de manifester la liberté d'indifférence.

149 La liberté absolue est la liberté qui est son propre motif, agissant volontiers contre la raison et la morale, afin de s'affirmer en tant que liberté.

150 Pour une description du phénomène, cf. O'REGAN, John Kevin and NOË, Alva. "A sensorimotor account of vision and visual consciousness", 5.10, p. 954. L'interprétation par les auteurs, de ce phénomène, est toutefois différente, mais non contradictoire avec l'hypothèse des « petites perceptions ».

ses tenants parlent parfois de « mémoire du corps » voire de « mémoire cellulaire », sans doute pour donner un vernis très sérieux à leurs spéculations mystiques…

Cette hypothèse permet à l'auteur des *Nouveaux Essais sur l'Entendement Humain* (toujours à travers le personnage de Théophile), de se livrer à quelques spéculations qui pourront intéresser les croyants aux vies antérieures et à la réincarnation[151] :

> « Or si ce passage des âmes était véritable, au moins de la manière possible que j'ai expliqué ci-dessus (mais qui ne paraît point vraisemblable), c'est-à-dire que les âmes, gardant des corps subtils, passassent tout d'un coup dans d'autres corps grossiers, le même individu subsisterait toujours dans Nestor, dans Socrate et dans quelque moderne, et il pourrait même faire connaître son identité à celui qui pénétrerait assez dans sa nature, à cause des impressions ou caractères qui y resteraient de tout ce que Nestor ou Socrate ont fait, et que quelque génie assez pénétrant y pourrait lire. Cependant si l'homme moderne n'avait point de moyen interne ou externe de connaître ce qu'il a été, ce serait quant à la morale comme s'il ne l'avait point été. Mais l'apparence est que rien ne se néglige dans le monde, par rapport même à la morale, parce que Dieu en est le monarque dont le gouvernement est parfait. Les âmes selon mes hypothèses ne sont point indifférentes à l'égard de quelque portion de matière que ce soit, comme il vous semble ; au contraire elles expriment originairement celles à qui elles sont et doivent être unies par ordre. Ainsi si elles passaient dans un nouveau corps grossier ou sensible, elles garderaient toujours l'expression de tout ce dont elles ont eu perception dans les vieux, et même il faudrait que le nouveau corps s'en ressentît, de sorte que la continuation individuelle aura toujours ses marques réelles. Mais quel qu'ait été notre état passé, l'effet qu'il laisse ne saurait nous être toujours apercevable. L'habile auteur de l'*Essai sur l'entendement*, dont vous aviez épousé les sentiments, avait remarqué (livre II, chap. *De l'identité*, § 27) qu'une partie de ses suppositions ou fictions du passage des âmes, prises pour possibles, est fondée sur ce qu'on regarde communément l'esprit, non seulement

[151] LEIBNIZ, Gottfried Wilhelm, *Les Nouveaux Essais sur l'Entendement Humain*, Livre II, Chapitre XXVII, § 14, p. 186.

> comme indépendant de la matière, mais aussi comme indifférent à toute sorte de matière. Mais j'espère que ce que je vous ai dit, Monsieur, sur ce sujet par-ci par-là, servira à éclairer ce doute, et à faire mieux connaître ce qui se peut naturellement. On voit par là comment les actions d'un ancien appartiendraient à un moderne qui aurait la même âme, quoiqu'il ne s'en aperçût pas. Mais si l'on venait à la connaître, il s'ensuivrait encore de plus une identité personnelle. Au reste une portion de matière qui passe d'un corps dans un autre ne fait point le même individu humain, ni ce qu'on appelle moi, mais c'est l'âme qui le fait. »

Notons qu'en dépit du fait que Leibniz attribue l'identité réelle à l'âme, cela n'en fait pas un partisan du critère psychologique. En effet, selon Locke, il est probable que l'âme est le siège de la conscience. Toutefois, pour lui, c'est la conscience, et non pas l'âme dont elle est le substrat, qui définit l'identité personnelle. Pour Locke, le siège de la conscience pourrait bien être le corps, ou une âme différente à chaque fois. Ce qui compte n'est pas l'hypostase mais la fonction que celle-ci accomplie. Leibniz, en revanche, tient l'âme en tant que telle pour siège de l'identité réelle. Or l'âme, tout comme le corps, est une hypostase.

On pourrait, si l'on veut, s'essayer à une petite typologie des diverses positions concernant l'identité personnelle et le problème âme-corps. On aurait ainsi :

- Des dualistes hypostatiques situant l'identité dans l'âme (ex. : Leibniz, etc.)

- Des dualistes hypostatiques situant l'identité dans le corps (ex. : St Thomas d'Aquin, avec certaines réserves concernant son statut de dualiste)

- Des monistes hypostatiques situant l'identité dans l'âme (ex. : Berkeley, etc.)

- Des monistes hypostatiques situant l'identité dans le corps (ex. : Stéphane Ferret, etc.)

- Des dualistes cognitionnistes situant la conscience dans l'âme : (ex. : Locke, etc.)

- Des dualistes cognitionnistes situant la conscience dans le corps

- Des monistes cognitionnistes situant la conscience dans l'âme

- Des monistes cognitionnistes situant la conscience dans le corps (les plus nombreux actuellement) : Shoemaker, Parfit, Perry, Lewis, Nozick, etc.

### 4. La meilleure image de l'âme humaine.

En descendant ainsi dans les profondeurs de la conscience, en vue de corriger le critère psychologique, nous sommes partis d'une connexité et continuité proprement psychologiques à une connexité et continuité qui ont quelque chose d'organique, de très proche du niveau biologique d'organisation. Bref, nous sommes passés au critère somatique de l'identité personnelle.

Un coup fatal paraît être porté au critère psychologique par Ronald de Sousa dans *Rational Homunculi*. L'auteur admet que[152] :

[152] La mise en forme a été ajoutée par nos soins.

> « Quand le Dr Jekyll se transforme en Mr Hyde, il se passe quelque chose d'étrange et de mystérieux. Sont-ils deux personnes occupant un même corps en alternance ? Mais voici quelque chose de plus étrange encore : le Dr Prodige et le Dr Miracle occupent, eux aussi, un corps à tour de rôle. La différence, c'est qu'ils sont aussi semblables que de vrais jumeaux ! Vous objectez : pourquoi dire, alors, qu'ils deviennent l'un l'autre ? Eh bien, pourquoi pas : si le Dr Jekyll peut devenir un homme aussi différent de lui que Hyde, il doit être d'autant plus facile pour Prodige de se changer en Miracle, qui lui ressemble trait pour trait.
>
> Il nous faut **un conflit** ou **une différence marquée** pour que nous puissions écarter notre idée **naturelle** qu'à ***un*** **corps** correspond **tout au plus** ***un*** **agent**. »[153]

Et si à un corps correspond tout au plus un agent, cela implique que Dr Jekyll et Mr Hyde sont une seule et même personne, en dépit de leur différence profonde de personnalité. Et si cela est vrai, alors le critère psychologique de l'identité personnelle, qui implique précisément le contraire, est donc erroné.

Cela n'implique toutefois pas de facto la justesse du critère somatique. Voyons comment se porte celui-ci face à un cas tel celui du Dr Jekyll et de Mr Hyde. Il est vrai que, à chaque moment, pour cet individu, tout se passerait comme s'il n'était qu'une seule personne. Si, lorsqu'il est Jekyll, il entend parler de ces frasques en tant que Hyde, il pensera qu'il s'agit de quelqu'un d'autre. Mais si quelqu'un le piste tous les jours, il saura quelle mésaventure surprenante lui arrive, et pourra l'identifier comme une seule et même personne. Or, cette reconnaissance sociale, comme l'avait justement remarqué Leibniz, est parfaitement légitime comme fondement, non seulement conventionnel mais aussi

153 DE SOUSA, Ronald. *Rational Homunculi.*

ontologique, de l'identité personnelle. Dans les *Nouveaux Essais sur l'Entendement Humain*[154], Leibniz (Théophile) écrit :

> « Je ne voudrais point dire non plus que l'identité personnelle et même le soi ne demeurent point en nous sous prétexte que je ne me souviens plus de rien de tout ce que j'ai fait alors. Il suffit pour trouver l'identité morale par soi-même qu'il y ait une moyenne liaison de consciosité d'un état voisin ou même un peu éloigné à l'autre, quand quelque saut ou intervalle oublié y serait mêlé. Ainsi, si une maladie avait fait une interruption de la continuité de la liaison de consciosité, en sorte que je ne susse point comment je serais devenu dans l'état présent, quoique je me souvinsse des choses plus éloignées, le témoignage des autres pourrait remplir le vide de ma réminiscence. On me pourrait même punir sur ce témoignage, si je venais de faire quelque mal de propos délibéré dans un intervalle, que j'eusse oublié un peu après par cette maladie. Et si je venais à oublier toutes les choses passées, et serais obligé de me laisser enseigner de nouveau jusqu'à mon nom et jusqu'à lire et écrire, je pourrais toujours apprendre des autres ma vie passée dans mon précédent état, comme j'ai gardé mes droits, sans qu'il soit nécessaire de me partager en deux personnes, et de me faire héritier de moi-même. (…) Ainsi, ne voulant point dire que l'identité personnelle ne s'étend pas plus loin que le souvenir, je dirais encore moins que le *soi* ou l'identité physique en dépend. L'identité réelle et personnelle se prouve le plus certainement qu'il se peut en matière de fait, par la réflexion présente et immédiate ; elle se prouve suffisamment pour l'ordinaire par notre souvenir d'intervalle ou par le témoignage conspirant des autres : mais si Dieu changeait extraordinairement l'identité réelle, la personne demeurerait, pourvu que l'homme conservât les apparences d'identité, tant les internes (c'est-à-dire la conscience) que les externes, comme celles qui consistent dans ce qui paraît aux autres. Ainsi, la conscience n'est pas le seul moyen de constituer l'identité personnelle, et le rapport d'autrui ou même d'autres marques y peuvent suppléer : mais il y a de la difficulté s'il se trouve contradiction entre ces diverses apparences. La conscience peut se taire comme dans l'oubli ; mais si elle disait bien clairement ce qui fut contraire aux autres apparences, on serait

[154] LEIBNIZ, Gottfried Wilhelm, *Les Nouveaux Essais sur l'Entendement Humain*, , Livre II, Chapitre XXVII, § 9, p. 183-184.

> embarrassé dans la décision et comme suspendu quelquefois entre deux possibilités, celle de l'erreur de notre souvenir et celle de quelque déception dans les apparences externes. »

Après tout, s'en tenir au critère psychologique, c'est accorder à un certain jugement (le sien propre, en tant qu'individu) une importance décisive relativement à la constitution ontologique de l'identité personnelle. Mais pourquoi privilégier un jugement en particulier, et lui accorder la toute-puissance au détriment de la multitude des autres ?[155] Pourquoi serions-nous le mieux placé pour dire qui nous sommes ? Ce préjugé, qui remonte à Descartes, a été battu en brèche par Wittgenstein, Ryle, la psychologie scientifique et une bonne partie de la philosophie analytique contemporaine, comme nous le verrons dans la partie suivante.

Pour eux, nous ne sommes pas le mieux placé pour dire qui nous sommes. Si cela est vrai de la personnalité, pourquoi cela serait-il faux de l'identité personnelle (au sens numérique)[156] ? Aussi, un critère psychologique « pluriel », « démocratique », « social », où cela sera l'ensemble des souvenirs de chaque individu à propos d'une personne particulière et non pas simplement les souvenirs propre à cette personne qui seront au fondement de son identité numérique, conduit naturellement au critère somatique. « Le corps humain est la meilleure image de l'âme humaine »[157].

---

155 Si la réalité ontologie a quelque chose à voir avec l'objectivité épistémologique, comme on peut être fondé à le penser en général, alors pourquoi serait-ce à la subjectivité pure que reviendrait le dernier mot, concernant la spécification de l'identité ontologique de la personne ?

156 Voir la partie consacrée à l'externalisme.

157 WITTGENSTEIN, Ludwig. *Recherches philosophiques*, II, iv, p. 254

# IV. De l'internalisme à l'externalisme

## *A. Introduction*

Le problème de l'identité personnelle se pose souvent comme celui de la recherche de propriétés que doit *avoir* un certain individu, ou de substance qu'il doit *être* pour qu'il soit une personne, une certaine personne, et une même personne au cours du temps. On présuppose alors que l'identité personnelle est déterminée par des critères internes, intrinsèques, à l'image des êtres naturels. Prenons par exemple une pièce de monnaie. La nature métallique de cette pièce (cuivre, or, argent, etc.) peut être déterminée en examinant la pièce elle-même, sans considération, semble-t-il, de quoi que ce soit d'autre. Par contre, sa valeur marchande n'est pas inscrite dans sa structure atomique, elle dépend de la situation socio-économique et géopolitique, des marchés financiers, des fluctuations de la bourse, etc. donc de facteurs externes, extrinsèques.

Se pourrait-il que l'identité personnelle dépende elle aussi de facteurs externes, au moins en partie ? Une telle conception, de prime abord choquante, aurait pourtant quelques avantages.

Si nous ressentons le besoin de résoudre un soi-disant mystère qui serait celui du fondement ontologique de l'identité personnelle, c'est bien parce que, d'un point de vue

cognitif, et dans les cas non pathologiques[158], il n'y a pas de problème véritablement insoluble qui se pose. Habituellement, on **re***connaît* immédiatement, de manière non-inférentielle, les personnes que l'on *connaît*. De même, et plus particulièrement, on ***se*** **re***connaît* normalement soi-même comme étant le sujet cognitif de ses propres souvenirs autobiographiques. Cela ne va pas de soi. Comment sait-on que nos souvenirs sont bien les nôtres et pas ceux de quelqu'un d'autre ? Dans un article paru en 2005[159], Ion Vezeanu nous propose une sorte de « renversement copernicien », pour venir à bout de cette difficulté. La connaissance, selon lui, n'est pas première par rapport à la reconnaissance, c'est la reconnaissance qui est première. La connaissance vient ensuite, qui la présuppose. L'impossibilité, pour un patient prosopagnosique, de reconnaître par exemple son propre visage, serait liée à une perturbation de sa capacité à individualiser les choses perçues. Cependant, on peut se demander si on est pas là en face d'une pétition de principe : puisque l'on n'arrive pas à justifier la reconnaissance en terme de connaissance, alors on fait l'inverse : on justifie la connaissance en terme de reconnaissance puis on postule que cette dernière n'a pas besoin d'explication…

---

158 La prosopagnosie est une pathologie curieuse dans laquelle un sujet devient incapable de reconnaître un être particulier comment étant cet être particulier, tout en demeurant capable de l'identifier comme étant tel type d'être, et de le décrire correctement. Le cas le plus fréquent touche la reconnaissance des personnes. Le sujet est incapable de reconnaître ses proches, mais sait qu'il s'agit d'êtres humains et est capable de les décrire normalement. En outre, il convient sans peine que ces personnes ressemblent beaucoup à ses proches, mais pour autant, il les perçoit comme des étrangers, voire, par conséquent, pour des imposteurs. Une hypothèse avancée est que c'est au niveau émotionnel que se joue ce déficit en reconnaissance. Nous associons un certain nombre de réactions émotionnelles à la perception de nos proches, et ce sont ces réactions émotionnelles qui nous permettraient de les reconnaître. Dans la prosopagnosie, ces réactions émotionnelles seraient altérées, d'où l'impossibilité de reconnaître les gens. Cela suggère que chez les sujets sains, il existe des réactions émotionnelles spécifiques pour chaque personne connue, sinon il ne serait pas possible de discriminer entre elles.

159 VEZEANU, Ion. « "Ce doit être moi puisque je suis ici !". Agnosie et reconnaissance ».

Il y a donc un fossé entre, d'une part, notre faculté de reconnaissance du même, en grande partie implicite et, normalement, d'une redoutable efficacité, et, d'autre part, notre grande difficulté à justifier cette efficacité en termes normatifs. Il ne semble pas y avoir de « recette miracle » facile à appliquer, qui permette, pour toute personne de dire : « Cette personne est X, a fait Y, etc. », et, pour tout un chacun, de se dire : « Je suis cette personne qui est X, a fait Y, etc. ». On pourrait se dire qu'après tout, la reconnaissance est un processus inconscient, ce qui est largement vrai, et en conclure un peu hâtivement qu'il est alors normal que l'on ne trouve pas « d'algorithme » de la reconnaissance du même, car un tel « algorithme » n'existe pas. Toutefois, que l'on ne suive pas de stratégie consciente lorsque l'on reconnaît quelqu'un n'implique pas qu'il n'y ait pas un processus rigoureusement déterministe à son fondement, seulement ce processus échappe à notre conscience.

Maintenant, pourquoi ne pas supposer que ce processus, quel qu'il soit, non seulement fonde notre capacité subjective à reconnaître le même, mais fonde objectivement cette *mêmeté* ? Ainsi, par exemple, « Jésus après sa Résurrection » serait la même personne que « Jésus sur la Croix » si (et peut-être seulement si) une majorité de personnes aux facultés cognitives normales (lui y compris) ayant connues le premier s'accordent ou s'accorderaient à reconnaître le second comme étant le même.

Et si tel n'est pas le cas, faudrait-il en conclure que les deux « Jésus » seraient numériquement distincts ? Pas nécessairement, mais alors il faut trouver un critère de la non-identité, qui serait, par exemple, qu'une majorité de personnes aux facultés cognitives normales (lui y compris) ayant connues le premier s'accordent ou s'accorderaient à exclure totalement le second comme étant le même. Mais quel serait le statut des ces êtres ne

satisfaisant à aucun des critères ? Peut-on concevoir une sorte de « No man's land » de l'identité numérique ? On peut aussi soutenir que : Y est la même personne que X **si et seulement si** une majorité de personnes aux facultés cognitives normales (Y y compris) ayant connues X s'accordent ou s'accorderaient à reconnaître Y comme étant X. Nous ne rentrerons pas, ici, dans ce débat.

Une image peut aider à comprendre comment des facteurs extrinsèques peuvent constituer une identité.

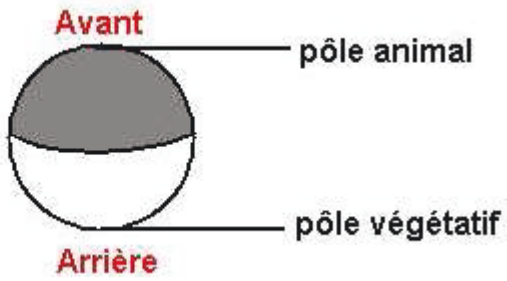
Imaginons un damier sur lequel sont disposés des pièces blanches qualitativement identiques, et des pièces noires qualitativement identiques. Deux personnes s'apprêtent à entamer une partie. Si nous nous absentons, les laissons jouer, et revenons en cours de partie, les pièces auront en grande partie bougées. Il y a bien sûr certaines règles selon lesquelles elles peuvent se déplacer. Néanmoins, le système n'est pas suffisamment déterministe pour qu'il soit possible de savoir précisément, dans tous les cas, quelle pièce est telle pièce. Au mieux, on pourra émettre quelques conjectures plus ou moins probables.

Considérons à présent les cases du damier. Nous avons des cases blanches, qualitativement identiques, et des cases noires, qualitativement identiques. A n'importe quel moment du jeu, il est possible de savoir précisément, dans tous les cas, quelle case est telle case. La différence n'est pas liée à des propriétés intrinsèques. Les cases d'une même couleur sont tout aussi semblables entre elles que les pièces d'une même couleur. La différence n'est pas liée au mouvement, possible dans le cas des pièces, apparemment impossible dans le cas des cases. En réalité, on peut toujours déplacer le damier. Et se faisant, on déplace bien les cases. Seulement voilà, on les déplace toutes en même temps et

de la même façon. Plus généralement, on applique au damier des modifications qui ne touchent pas sa structure topologique. Tant que celle-ci est préservée, l'identification numérique des cases est possible. En effet, nous remarquons que les cases d'un damier qui serait en caoutchouc mou seraient tout aussi parfaitement identifiables, à n'importe quel moment, que les cases d'un damier normal (en bois ou en plastique par exemple). Qu'est-ce que la structure topologique du damier, par rapport à chaque case ? Ce n'est pas une propriété intrinsèque des cases, mais une propriété du damier lui-même, donc une propriété extrinsèque aux cases. C'est donc bien une propriété extrinsèque qui, dans le cas des cases du damier, permet de les réidentifier d'un moment à l'autre. Du point de vue des cases, ces propriétés se présentent comme un certain nombre de relations constantes qu'entretient chaque case avec toutes les autres.

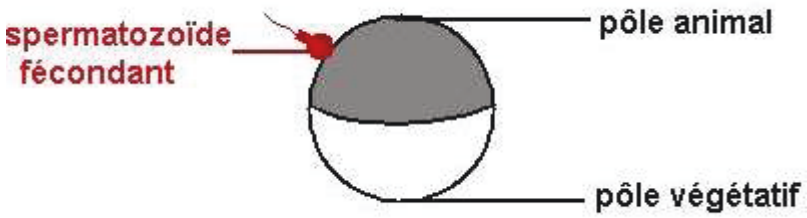
A-t-on, chez les personnes, des relations constantes du même genre, que chacune entretient avec toutes les autres ? Dans le cas du damier, les relations constantes sont de type spatial, structural. Il ne semble pas que des relations constantes de ce type puissent caractériser les personnes les unes par rapport aux autres. Une meilleure image serait peut-être celle d'un organisme dont chaque personne serait une cellule. Un organisme n'est pas une structure figée. Son unité est dynamique, fonctionnelle, complexe. L'embryologie nous montre un exemple de cette détermination complexe d'identité par des facteurs externes, fonctionnels et dynamiques : *les champs morphogénétiques*. Ceux-ci, d'après les dernières recherches des embryologues, s'incarnent dans différents gradients de concentration de molécules appelées « morphogènes » (généralement des protéines ou des ARNm), à travers l'ovule, le zygote et/ou l'embryon. Deux mécanismes peuvent expliquer la formation de ces gradients selon le moment où ils se forment :

- **Formation du gradient au stade ovule/zygote :** Les molécules morphogènes se répartissent inégalement par diffusion à travers la cellule. Ensuite, lors de la segmentation du zygote, les cellules héritent, en fonction de leur position dans l'embryon en formation, d'une fraction de cytoplasme qui va contenir, par conséquent, plus ou moins de molécules morphogènes. On observe ce mécanisme chez des amphibiens comme le crapaud xénope (voir schémas ci-dessous[160]).

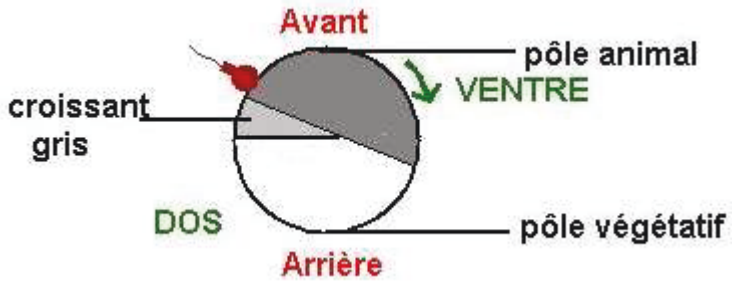



L'ovule de xénope est inégalement réparti en vitellus, définissant un pôle dit « animal », pauvre en vitellus et contenant le noyau, opposé à un pôle dit « végétatif », riche en vitellus. Dès ce stade, une première détermination s'effectue : le pôle animal donnera la partie antérieure de l'animal, tandis que le pôle végétatif donnera la partie postérieure.

[160] Ces schémas sont issus de DELARUE, Michel, « "De l'oeuf à la grenouille". Les Etapes du Développement d'un Amphibien : Le xenope (Xenopus laevis) » [en ligne].




Ensuite, vient la fécondation. L'emplacement d'entrée du spermatozoïde à son importance pour la suite du processus.

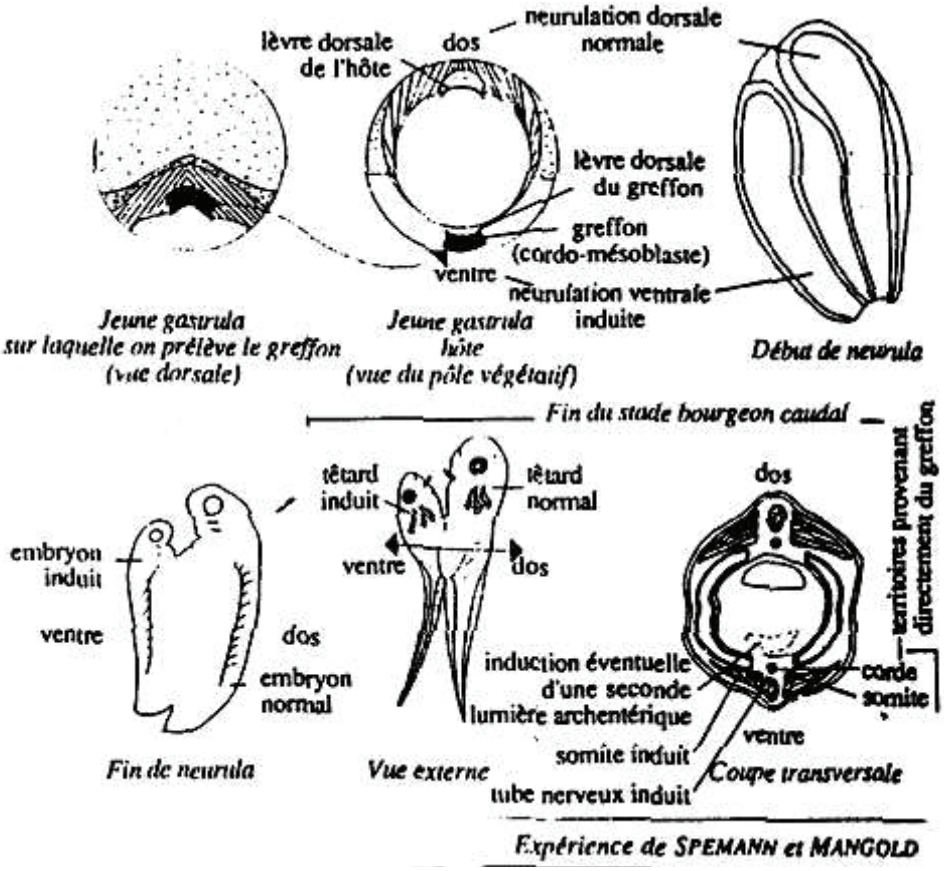



En effet, cela entraîne, à ce niveau, une rotation de 30° du cytoplasme cortical qui donne naissance au croissant gris. On a alors ici une deuxième détermination puisque ce croissant gris donnera le futur dos de l'animal.

- **Formation du gradient au stade embryon :** Les molécules morphogènes se répartissent inégalement par diffusion à travers les cellules de l'embryon (par endocytose ou par des canaux protéiques) au stade gastrula, à partir d'un tissu déjà différencié, nommé « centre organisateur ». Ce phénomène d'induction embryonnaire (car le centre organisateur induit la détermination des cellules totipotentes adjacentes) a été mis en évidence par Hilde

Mangold et Hans Spemann en 1924[161], ce qui valut le prix Nobel à ce dernier en 1935. On l'observe aussi chez les amphibiens, au stade gastrula.

L'expérience menée par Mangold et Spemann peut être schématisée comme suit :




On greffe une région cordo-mésoblastique dans la partie ventrale d'une jeune gastrula hôte. Quelques temps après, on observe une double neurulation, puis la formation d'un embryon double. Une coupe transversale montre qu'à part une des deux cordes, provenant directement du greffon, tous les tissus proviennent exclusivement de l'embryon

[161] SPEMANN, Hans und MANGOLD, Hilde. „Über Induktion von Embryonalanlagendurch Implantation artfremder Organisatoren".

hôte. L'embryon dont la corde provient du greffon, et qui n'aurait pas dû apparaître normalement, a donc été induit sur les territoires de l'hôte par le greffon.

Quel que soit le scénario de mise en place du gradient, le principe est le même : les cellules ont la capacité de détecter, par ces gradients de concentration, leur position dans l'embryon, et de s'auto-déterminer en conséquence.

Peut-on transposer ce modèle de la détermination de l'identité cellulaire lors du développement embryonnaire à la détermination de l'identité personnelle ?

Une première objection pourrait être la suivante : la détermination cellulaire, même si elle dépend d'un gradient chimique, se concrétise par une modification des cellules. Cette modification n'est peut-être pas très visible tout de suite, mais pourtant elle est bien réelle : dans les cellules embryonnaires, au moment de la détermination, des modifications opèrent au niveau du noyau, empêchant l'expression d'autres gènes que ceux qu'elles exprimeront une fois différenciées. Cela revient à dire qu'ultimement, la détermination cellulaire repose sur des facteurs internes.

Qu'il y ait des facteurs internes, nous ne le nions pas. Ce que nous voulons savoir, c'est si ces facteurs doivent être considérés comme ontologiquement prééminents et suffisants.

Une deuxième objection pourrait être la suivante : les facteurs externes, dans la détermination cellulaire, sont des molécules. Les molécules sont des entités réelles. A l'échelle de l'humanité, il n'y a guère de médiateurs chimiques entre individus. Il y a bien les

phéromones, mais ce qui semble le plus déterminant passe par de l'information visuelle et sonore essentiellement, et ce qui est le plus déterminant encore, dans cette information, c'est non pas la nature de son support, mais la forme. Ainsi, qu'un texte soit rédigé sur une plaquette d'argile, un parchemin, une feuille A4 90g imprimée, etc. ce qui va le plus compter pour nous, avoir le plus d'effet, cela va être le contenu linguistique, codé symboliquement sur ce support, et non le support lui-même. C'est par conséquent quelque chose de très abstrait, de très symbolique, consciemment traité, qui va nous influencer majoritairement, et donc quelque chose dont on peut suspecter qu'il n'a pas une réalité aussi forte que le support sur lequel il s'inscrit. Mais sur quoi se base cette conception matérialiste ontologique qui ferait des seuls éléments matériels, en tant que tels, sans considération de leur agencement et de leur dynamique, les dépositaires de toute réalité ultime ? A priori, il ne semble pas plus insensé de dire que c'est la forme qui détermine la matière que d'affirmer l'inverse.

Une troisième objection à cette hypothèse s'appuie sur une seconde dichotomie, entre *réel* et *conventionnel*, qui est souvent alignée sur la première (entre critères internes/intrinsèques et critères externes/extrinsèques). Une longue tradition de pensée nous conduit à considérer les propriétés intrinsèques comme seules authentiques, comme ontologiquement premières, donc comme seules proprement réelles. Le « Tout » n'existe pas en soi, il n'est que la résultante du fonctionnement coopératif de ses parties. On peut bien sûr parler des propriétés extrinsèques comme si elles avaient une véritable profondeur ontologique, comme si elles avaient du « corps » et de la « consistance », une efficience causale, une réalité substantielle. Mais ce n'est que pure « façon de parler » commode, simple convention de langage qui ne reflète pas la structure réelle du monde. Selon ce point de vue, que l'on peut qualifier de nominaliste, des problèmes fameux, comme les paradoxes

sorites au sujet des tas de sable à un seul grain, des nains de 2 m et autres chauves fort bien chevelus, ne sont que des problèmes fumeux. Ils sont artificiellement maintenus en vie par quelques philosophes aveuglés par la « magie des mots ». Ces philosophes ne voient pas que des notions comme « tas de sable », « nain », « chauve », etc. ne dépeignent en fait aucune réalité métaphysique sur laquelle il pourrait être intéressant ou utile de discourir. Ils ne se rendent pas compte que ce sont simplement des mots-outils, sans référent clairement identifiable, et dont le sens s'inscrit dans et se réduit à une pratique, un usage, dans certaines situations et pour répondre à certains besoins bien définis des gens. D'ailleurs, pour les plus « extrêmes », parmi les nominalistes, que ces notions engendrent ce genre de pseudo-problèmes est un argument fort en faveur de leur irréalité. Ainsi, en matière d'identité personnelle, Derek Parfit[162] adopte une position antiréaliste face au constat qu'aucune réponse entièrement satisfaisante ne peut être apportée aux problèmes qu'elle pose. On n'arrive pas clairement à définir ce qui fait de telle personne cette personne en particulier tout au long de sa carrière ? C'est simple : c'est donc qu'il n'y a rien à définir !

Pour d'autres, c'est aller trop loin. Que les « tas de sables » n'existent pas réellement, cela ne dérange personne. Que les personnes n'existent pas réellement, cela prend à revers une autre tradition de pensée, à laquelle on tient en général beaucoup plus. Par exemple, l'existence réelle des personnes est au fondement de notre morale. Aussi, remettre en question cette première, n'est-ce pas remettre en question cette dernière ? Si Derek Parfit s'en accommode bien, on ne peut pas dire que la morale qu'il préconise soit exempte de problèmes et de questionnements. Elle ne va pas de soi. Par exemple, la notion de souffrir un peu maintenant pour un plus grand bien plus tard devient totalement obscure dans sa morale.

---

[162] PARFIT, Derek. *Reasons and Persons*.

On peut se demander, toutefois, si l'association fréquente, d'une part, entre les propriétés intrinsèques et les propriétés réelles, d'autre part, entre les propriétés extrinsèques et les propriétés conventionnelles, est bien justifiée.

Dans la suite de cette partie, nous nous attacherons à renverser conjointement toutes ces idées selon lesquelles :

- Les facteurs internes doivent être considérés comme ontologiquement prééminents sur les facteurs externes et suffisants

- La matière (les éléments matériels) constitue la seule réalité ; elle détermine la forme (structure, fonction, dynamique, etc.) qui en soi n'a aucune efficacité causale

- Ce qui est intrinsèque est réel, ce qui est extrinsèque est purement affaire de convention.

## *B. Perspective épistémologique*

### 1. L'argument du langage privé et ses variantes

Enfin, il nous faut mentionner une critique importante, reprise à différentes sauces par plusieurs philosophes contemporains de la tradition analytique, et inspirée par les remarques de Wittgenstein présentes aux paragraphes 243 à 315 des *Recherches*

*Philosophiques*[163]. Il s'agit d'un argument contre la possibilité d'un langage privé. Cette critique wittgensteinienne, qui ruine toute possibilité de fonder l'édifice de la connaissance sur des bases subjectives, s'attache à souligner l'incohérence conceptuelle inhérente à la thèse, souvent défendue par le cartésianisme et le post-cartésianisme, selon laquelle l'introspection serait comparable à une perception, par un sens ou quasi-sens interne, d'états mentaux essentiellement privés. L'argument montre que si les états mentaux sont essentiellement privés, ils ne peuvent être dit perçus et a fortiori connus par le sujet qui les expérimente et que s'ils sont perçus et/ou connus par lui, alors ils ne peuvent être essentiellement privés. Avant d'expliciter cet argument, voyons d'abord en quoi consiste exactement la double thèse selon laquelle, d'une part, l'introspection serait une sorte de perception de ses propres états mentaux (ce qu'il convient d'appeler le modèle perceptif de l'introspection), et d'autre part, les états mentaux seraient essentiellement privés.

A la suite de Descartes notamment, on a longtemps cru que quelqu'un observant et décrivant ses propres états mentaux était exactement dans la même situation, par rapport à ces derniers, que lorsqu'il observe et décrit un objet extérieur, par rapport à cet objet. C'est ce qu'on appelle le modèle perceptif de l'introspection. Observer ses états mentaux, ce n'est ni plus ni moins la même chose qu'observer une pomme ou une chaise.

Quant à dire que les états mentaux sont privés, cela revient à dire qu'ils ne sont accessibles qu'au sujet qui en fait l'expérience. Autrement dit, une approche objective des états mentaux serait impossible. C'est ce que soutient Nagel[164] qui donne l'exemple de l'écholocalisation chez la chauve-souris. On aurait beau connaître très précisément toute la

163 WITTGENSTEIN, Ludwig. *Recherches Philosophiques*, § 243-315, p. 135-155.

164 NAGEL, Thomas, "What it is like to be a Bat?".

physiologie de cet animal, cela ne nous donnerait aucune idée de l'effet que cela fait d'expérimenter le monde à travers la représentation que s'en fait la chauve-souris grâce à son système d'écholocalisation. L'expérience subjective qualitative, demeurerait donc à jamais confinée dans une intériorité inaccessible à la science objective.

Cette double thèse d'un accès (1) perceptif à (2) des états mentaux strictement privés n'est pas sans poser problème. Le savoir suppose un langage, dans lequel sera formulé un certain nombre de propositions – vraies ou fausses – sur le réel. Ces propositions sont formées de mots définis par d'autres propositions ou par désignation d'objets du monde réel (définition déictique ou ostensive). Par conséquent, si l'introspection nous donne accès à un savoir, on doit s'attendre à ce qu'il existe un certain langage dans lequel sont formulées les propositions – vraies ou fausses – portant sur les états mentaux. Ces propositions seraient formées de mots définis par d'autres propositions ou par désignation d'états mentaux internes. Puisque notre conscience n'a pas de doigt pour désigner un état mental privé, la cérémonie établissant la règle d'utilisation d'un nom se rapportant à une sensation prendrait la forme d'une attention soutenue à un des aspects de notre paysage mental, pendant que l'on se dirait à soi-même « J'appelle cette sensation S ! ». On rencontre ici un premier problème. Lorsque je désigne un objet public puis le nomme, je ne suis pas prisonnier de ma subjectivité, d'autres personnes peuvent prendre part à cette cérémonie, observer l'objet que je pointe, et discuter avec moi de son nom – peut-être en a-t-il déjà un que j'ignore et que l'on pourra m'apprendre ? – me demander des précisions si ma désignation était trop vague – Quine[165] a montré que le langage comportait toujours une ambiguïté – ou encore adopter ce nom pour désigner le même objet public. Cela ne peut évidemment pas être le cas si je désigne et nomme une sensation. Personne ne comprendra

[165] QUINE, Willard Van Orman. *La poursuite de la vérité*.

clairement de quoi je parle et, à supposer que d'autres personnes expérimentent parfois la même sensation que moi et qu'ils leur prennent l'envie de lui attribuer aussi un nom, il n'est pas dit que ce sera le même. En croyant expérimenter le même état phénoménal que moi, quelqu'un pourra vouloir désigner le sien par le même nom, alors qu'il est possible qu'il n'expérimente en fait pas tout à fait voire pas du tout la même chose. Aucun moyen de comparaison n'existe !

Comme le fait remarquer Wittgenstein[166] :

> « 293. Si je dis de moi-même que je sais seulement à partir de mon propre cas ce que signifie le mot « douleur », — ne faut-il pas que je *le* dise aussi des autres ? Et comment puis-je donc généraliser ce *seul* cas avec tant de désinvolture ?
>
> Eh bien, tout le monde vient me dire qu'il ne sait qu'à partir de son propre cas ce qu'est la douleur ! —— Supposons que chacun possède une boîte contenant ce que nous appellerons un « scarabée ». Personne ne pourrait jamais regarder dans la boîte des autres ; et chacun dirait qu'il ne sait ce qu'est un scarabée que parce qu'il a regardé le *sien*. — En ce cas, il se pourrait bien que nous ayons chacun, dans notre boîte, une chose différente. On pourrait même imaginer que la chose en question changerait sans cesse. — Mais qu'en serait-il si le mot « scarabée » avait néanmoins un usage chez ces gens-là ? — Cet usage ne consisterait pas à désigner une chose. La chose dans la boîte ne fait absolument pas partie du jeu de langage, pas même comme un *quelque chose* : car la boîte pourrait aussi bien être vide. — Non, cette chose dans la boîte peut être entièrement « supprimée » ; quelle qu'elle soit, elle s'annule.
>
> Cela veut dire : Si l'on construit la grammaire de l'expression de la sensation sur le modèle de « l'objet et sa désignation », l'objet perd toute pertinence et n'est plus pris en considération. »

[166] WITTGENSTEIN, Ludwig. *Recherches Philosophiques*, § 293, p. 151.

Si un terme du langage mental a un sens, ce sera donc un sens privé, auquel je serai le seul à avoir accès, étant par hypothèse la seule personne à savoir quel effet cela fait d'être moi. Ce terme ne pourra donc pas servir à désigner un type de sensation qui pourrait être commun à plusieurs personnes et permettre un certain accord intersubjectif malgré le caractère privé de son référent. On aurait sans doute l'impression de tous parler de la même chose en employant le même terme et peut-être aussi de ne pas parler de la même chose en employant un terme différent, mais en réalité, il n'y a aucun moyen de le savoir et cette impression est donc tout bonnement illusoire. Nous allons voir maintenant que nous nous illusionnons aussi si nous espérons accéder malgré tout, individuellement, à un sens privé. En effet, y a-t-il un sens à parler d'un sens qui ne soit que pour moi, qui soit confiné dans une intériorité cachée à tout le monde à part moi ? Pour qu'un langage ait un sens, il faut qu'il soit utilisé correctement. Or, pour que l'on puisse dire qu'un langage est utilisé correctement, il faut que demeure possible son usage de manière erroné. Pour cela, il faut une norme qui permette de discriminer les usages corrects et fautifs du langage. Or une norme est nécessairement une construction sociale et le respect de cette norme est soumis à un contrôle intersubjectif. S'il existait des normes internes, qui serait le garant de leur respect effectif ? Comment distinguer les cas où j'applique correctement la règle d'utilisation d'un terme du langage phénoménal d'un cas où je crois seulement, mais à tort, l'appliquer correctement ? Je ne peux pas à la fois me tromper dans l'usage d'un terme du langage privé et avoir conscience que je me trompe. Bien sûr, après coup, je peux toujours me dire « Je me suis trompé ». Mais ce jugement est lui-même inscrit dans la subjectivité. Une norme interne peut fluctuer sans cesse – même inconsciemment et d'ailleurs le plus souvent inconsciemment – car elle ne dépend que du sujet. Une norme externe est intersubjective. Elle ne peut changer que par consensus. On aimerait pouvoir dire ici, comme le mentionne Wittgenstein, que tout ce qui va me sembler correct sera correct, mais

cela veut seulement dire que le terme « correct » n'a aucun sens dans ce contexte. Un jeu où chaque joueur a le pouvoir de modifier les règles à sa guise n'est plus un jeu !

Nous avons donc montré qu'il est impossible de fixer une signification à un terme d'un hypothétique langage mental, car ce terme ne serait compréhensible de personne d'autre que moi et surtout, étant juge et partie, je n'aurais aucun moyen de savoir si j'emploie ce terme correctement, à chaque fois que je l'emploie. L'argument, on le voit, ne se limite pas à dire qu'une définition ostensive d'une sensation privée établit une règle que je ne peux pas suivre, il dit en outre qu'elle n'établit aucune règle du tout.

Nous pouvons ici rappeler une « intuition pump » intéressante de Dennett[167], qui va dans le même sens que la critique wittgensteinienne :

Chase et Sanborn sont deux goûteurs de café chez Maxwell depuis 6 ans. Leur travail consiste a vérifier que l'arôme des cafés Maxwell reste constant, année après année. Tous deux s'accordent à dire que le café Maxwell n'a pas changé depuis tout ce temps. Toutefois, alors qu'ils aimaient ce café autrefois, lorsqu'ils ont commencé à travailler, ils ne l'aiment plus maintenant. Cependant, leurs raisons sont différentes :

(A) Chase pense que le café a toujours le même arôme pour lui mais que ses goûts ont changé, qu'il est devenu un buveur de café plus sophistiqué.

[167] DENNETT, Daniel Clement. *Quining qualia.*

(B) Sanborn quant à lui, pense que ses goûts n'ont pas changé mais que c'est sa perception de l'arôme du café qui a changé. Il pense que quelque chose doit dysfonctionner quelque part dans son système perceptif. Toutefois, il admet que l'arôme du café Maxwell en lui-même n'a pas dû changer puisque tous les autres goûteurs sont d'avis qu'il est toujours le même.

Qu'en est-il réellement ? Envisageons tout d'abord le cas de Chase. Il apparaît de prime abord 3 possibilités :

(a) Les qualia de Chase induits par l'arôme du café sont, comme il le pense, restés constants, mais sa réaction à ces qualia a changé.

(b) Chase se trompe en pensant que ses qualia n'ont pas bougés. Ceux-ci se sont modifiés graduellement et imperceptiblement au fil des ans tandis que ses critères de jugement, contrairement à ses dires, sont restés constants.

(c) La vérité se trouve quelque part entre (a) et (b). Ses qualia ont un peu changé et ses goûts aussi.

Le cas de Sanborn implique 3 possibilités semblables :

(a) Sanborn a raison. Ses qualia ont changé, en raison d'un dysfonctionnement de son système perceptif, mais ses goûts sont toujours les mêmes qu'avant.

(b) Sanborn se trompe. Ses critères esthétiques ont changé à son insu mais ses qualia sont toujours identiques.

(c) La vérité se trouve quelque part entre (a) et (b). Ses qualia ont un peu changé et ses goûts aussi.

Dans cette expérience de pensée, ce sont les goûts ou jugements des deux protagonistes qui jouent le rôle de norme interne et les qualia correspondants à la perception de l'arôme du café qui sont les états mentaux que l'on prétend connaître. L'exemple montre que la critique wittgensteinienne est transposable à des domaines de l'accès au contenu de notre conscience qui n'ont qu'un rapport indirect au langage[168]. Il apparaît donc impossible, contrairement à ce qu'on croit souvent, de connaître ses propres qualia – si on conçoit ceux-ci comme privés – que ce soit conceptuellement, par le biais du

[168] Bien sûr, le langage n'est pas entièrement absent. Il sert à formuler des propositions telles que « L'arôme du café est toujours le même pour moi mais ce sont mes goûts qui ont changé » ou « J'ai toujours les mêmes goûts qu'avant mais c'est ma perception de l'arôme du café qui a changée ». La perception de l'arôme du café – en plus d'être évaluée par nos jugements de goût – et la perception (illusoire ?) de l'évolution de cette perception sont également associées à un mot ou une expression verbale. Que cette association jouisse d'une certaine stabilité dans le temps et l'on pourra la qualifier de norme. De même, nos jugements de goût et la perception (illusoire ?) de leur évolution sont également associés à un mot ou une expression verbale. Que cette association jouisse d'une certaine stabilité dans le temps et, là aussi, on parlera de norme. Notons en passant qu'une norme peut être normalisée par une autre norme et tient donc lieu, pour cette dernière, de fait. Le problème soulevé par l'expérience de pensée de Dennett se transpose sans peine à ce niveau linguistique, et rejoint de fait la critique de Wittgenstein.

langage, ou même par l'entremise d'une représentation plus « imagée » basée sur l'observation de nos réactions émotionnelles.

## 2. La connaissance par contact direct : réalité et limites

Mais que penser d'une connaissance plus directe, une *« knowledge by acquaintance »* pour reprendre l'expression de Betrand Russell ?

Avant tout, il faut voir ce que l'on doit entendre par là. L'expression de Bertrand Russell est difficilement traduisible en français[169], sinon en insistant sur une distinction entre une connaissance au sens strict et le savoir, comme connaissance par description (*knowledge by description* chez Russell) ou connaissance conceptuelle. Cette distinction se fait plus saillante lorsque l'on pense à certaines expressions, dans la langue de Molière, où les termes de « connaissance » et de « savoir » sont utilisés. Ainsi, *on connaît Pierre* mais *on ne sait pas (que) Pierre*, on peut *savoir que Pierre est quelqu'un de très grand, qui parle beaucoup, qui a deux voitures, qui mange comme trois, etc.* ce qui nous donne une idée du **type** de personne qu'est Pierre. Mais quand on dit qu'*on connaît Pierre*, cela veut dire que l'on connaît personnellement Pierre, un certain Pierre, et pas un autre qui lui ressemblerait comme deux gouttes d'eau. Connaître Pierre, c'est connaître un certain **token**, une certaine occurrence singulière de la classe des personnes, c'est-à-dire l'identifier, le re*connaître*. En revanche on ne dira pas qu'on *connaît que Pierre est quelqu'un de très grand, qui parle beaucoup, qui a deux voitures, qui mange comme trois, etc.* Connaître met l'accent sur le sujet, pas sur ses attributs, qui

[169] Le terme français d'accointance, parfois utilisé comme traduction d'*acquaintance*, ne nous paraît pas des plus heureux, puisque le mot est défini comme suit par le Petit Larousse : **Accointances** n.f. pl. (du lat. *accognitus*, connu). *Péjor.* Relations, fréquentations jugées peu recommandables.

peuvent d'ailleurs être partagés par d'autres sujets. Savoir, au contraire, met l'accent sur les attributs, pas sur le sujet qui les porte, qui peut d'ailleurs ne pas être unique. Enfin, la connaissance semble plutôt porter sur les essences, c'est-à-dire sur les définitions, les conditions nécessaires et suffisantes pour qu'un être soit l'être qu'il est, indépendamment de la question de son existence réelle. Le savoir, au contraire, a trait au monde réel, il exprime un jugement sur les choses telles qu'elles sont en réalité. La connaissance est donc antérieure au savoir, puisqu'il faut déjà s'être entendu sur la signification de tel ou tel concept, avant de pouvoir construire des énoncés synthétiques portant sur ces concepts.

Il semble que certaines expressions telles que *« Je connais les mammifères »* contredisent les règles d'utilisation des termes « connaître » et « savoir » que nous avons mis en évidence ci-dessus. Les mammifères ne sont-ils pas une classe, et donc un type d'animaux et non pas une occurrence (token) de la classe des animaux ? Oui et non. L'opposition type-token n'est pas toujours absolue. Lorsque l'on s'intéresse à une classe, on peut bien entendu la voir comme un type qui va instancier différents token mais on peut aussi l'envisager comme un certain token instancié par une classe plus vaste. Les mammifères sont ainsi une occurrence particulière de la classe des animaux, parmi d'autres occurrences de même niveau tels les oiseaux ou les reptiles. Il n'existe qu'une seule classe des mammifères, et cette classe aurait pu ne pas exister ou disparaître. Enfin, la classe des mammifères peut être identifiée assez précisément et reconnue. On peut, notamment, la circonscrire à travers des caractéristiques physiologiques et anatomiques, en retracer les origines et l'histoire, en délimiter la répartition.

Tout cela nous donne une première notion de la connaissance par opposition au savoir. On précisera[170] que dans l'esprit de Bertrand Russell, la connaissance par *acquaintance* est une connaissance directe : *« Nous dirons que nous avons l'expérience directe (acquaintance) d'une chose, quand elle est là directement devant nous, que nous en avons conscience, sans l'intermédiaire d'aucun processus d'inférence ou de quelque connaissance de vérité que ce soit »*[171]. Son objectif est fondationaliste : *« Existe-t-il au monde une connaissance si certaine qu'aucun homme raisonnable ne puisse mettre en doute ? »*[172]. La réponse de Bertrand Russell est positive : les *sense data* (les données sensorielles) sont ce type d'objet que l'on connaît de manière absolument certaine. A ce niveau là de conceptualisation, nous nous détachons du langage ordinaire, puisque le concept de connaissance que nous avons évoqué plus haut, s'il s'accommode bien avec l'idée de connaissance directe – lorsqu'on connaît Pierre, l'Amérique, la Neuvième Symphonie de Beethoven, etc. l'idée sous-jacente est aussi qu'on les reconnaît d'emblée, dans leur unité, comme un tout, sans avoir à pratiquer une inférence, tandis que lorsque l'on sait que Pierre mesure 2m10, cela implique de se représenter cette mesure et de l'associer à une représentation de Pierre – s'accommode beaucoup moins avec le projet fondationaliste de Bertrand Russell, puisque connaître Pierre, ça ne veut pas dire qu'il soit impossible de prendre quelqu'un d'autre pour lui ou de ne pas le reconnaître (par exemple si cela fait plusieurs années qu'on ne l'a pas vu et qu'il a beaucoup vieilli).

---

170 Pour cette précision et les suivantes de ce paragraphe, nous nous sommes largement inspiré de l'article de CHAPUIS, Delphine : *Les sense data jouent-ils un rôle dans la connaissance ?* [en ligne] qui fait une synthèse intéressante de ces questions.

171 RUSSELL, Bertrand. *Problèmes de philosophie*, p. 69

172 Première phrase de Russell, Bertrand. *Problèmes de philosophie.*

On peut douter, avec Schlick[173], que la connaissance par *acquaintance* soit belle et bien une connaissance. En effet, pour Schlick, connaître un fait, c'est en connaître la structure. Que faut-il entendre par là ? Tout d'abord, la connaissance suppose un acte de reconnaissance et s'oppose donc à la passivité contemplative de l'intuition. Ensuite, connaître un objet, c'est l'identifier comme tel et tel et le situer par rapport aux objets précédemment connus donc l'insérer dans un vaste réseau de propositions connectées les unes aux autres de manière cohérente. Autrement dit, une connaissance ne peut être isolée. Or, dans l'*acquaintance*, on assiste à une absolutisation du sujet et de l'instant présent, qui s'oppose à la connaissance comme processus historique individuel et collectif. On rejoint ici les critiques que nous avons déjà vues.

On peut également s'interroger sur la pertinence de l'acquaintance dans le fondement de la connaissance (en tant que savoir conceptuel et descriptif). Comment un processus par nature essentiellement privé pourrait-il servir de socle à une connaissance qui elle s'inscrit nécessairement dans un certain langage intersubjectif et porte sur des objets publics ? La critique du langage privé frappe fort. Pourtant il semble contre-intuitif de considérer que le donné sensible ne joue absolument aucun rôle dans la connaissance. Sinon, à quoi servirait-il ? Du reste, on a du mal à imaginer qu'un être qui, par hypothèse, n'aurait aucune vie subjective puisse être dit connaître quoi que ce soit[174]. Enfin, la connaissance ne repose-t-elle pas essentiellement sur l'observation ? Et qu'est-ce qu'observer sinon prendre conscience d'un objet à travers l'effet que celui-ci produit sur notre système sensoriel ? Le langage lui-même, pour être compris, doit d'abord être perçu, et cette perception ne revient-elle pas à l'appréhender à travers l'effet subjectif que la parole ou l'écrit réalise en

---

173 SCHLICK, Moritz, "Is there intuitive knowledge?"

174 Cf. l'expérience de pensée de la chambre chinoise (SEARLE, Richard John, *La redécouverte de l'esprit*)

nous ? C'est un problème sur lequel buta Schlick qui partageait les vues fondationalistes de Bertrand Russell. Schlick, ainsi d'ailleurs que Rudolf Carnap[175], entrevoyait la possibilité d'un langage privé sous la forme d'énoncés protocolaires. Même si ce langage n'avait pas réellement de sens (cf. la critique wittgensteinienne), il devait permettre de comprendre comment on peut passer, à titre personnel, du donné sensible à la connaissance et réciproquement. Pour Otto Von Neurath[176], une telle tentative est boiteuse. Il critiqua vivement toute prétention à fonder le sens et la connaissance dans l'intériorité de l'expérience privée. Il se base essentiellement sur une critique du langage privé, proche de celle que nous avons déjà évoquée.

En l'occurrence, un langage qui serait vraiment phénoménal – à supposer qu'il soit possible – ne devrait pas comporter des termes du langage naturel comme « rouge », « rond », « froid », « douloureux ». En effet, le langage naturel est intersubjectif, c'est-à-dire que la signification de ses termes ne se trouve pas dans l'intériorité d'une expérience subjective privée, qui par définition peut être très différente d'un individu à l'autre sans qu'il soit possible de détecter cette différence[177]. Ainsi, utiliser de tels mots pour qualifier les expériences privées, c'est effectuer un glissement sémantique : passer d'une désignation d'objets publics, quel que soit l'effet subjectif que ces objets engendrent chez les gens, à une désignation d'aspects privés engendrés par l'expérience de ces objets publics.

Mais peut-on vraiment désigner autre chose que ses expériences mentales ? Peut-on vraiment connaître autre chose que ses sense data ? Bref, Neurath ne va-t-il pas trop loin en

---

175 Cf. CARNAP, Rudolf, *La construction logique du monde*.

176 VON NEURATH, Otto. "Radical Physicalism and the Real World".

177 Cf. l'expérience du spectre inversé

déniant tout rôle à nos sensations privées ? Est-il même possible de prendre au sérieux toutes ces critiques du langage privé que nous avons vu jusqu'à présent ? Bien sûr, théoriquement, cela a l'air valable, mais est-ce que c'est encore défendable en pratique ? Les gens qui confient leurs sentiments à leurs proches savent bien de quoi ils parlent et se font généralement assez bien comprendre, et quel wittgensteinien ne s'est-il pas déjà parlé à lui-même en pensée sans avoir l'impression d'être insensé ?

Il ne faut pas oublier que la critique du langage privé vaut surtout pour un langage qui serait **essentiellement** privé, ce qui est un préjugé cartésien et post-cartésien. En effet, dans le cartésianisme, les états mentaux sont des propriétés intrinsèques d'un Ego simple, indivisible et immatériel. Une telle entité, dans un vocabulaire cette fois-ci leibnizien, est donc une monade. Elle n'a « ni porte ni fenêtre »[178].Or, quand bien même nous aurions des qualia, sense data et autres sensibilia, pourquoi les concevrions-nous comme essentiellement privés ? Nous ne sommes pas obligés de souscrire au modèle de l'Ego cartésien, et nous avons d'ailleurs vu précédemment d'excellentes raisons de ne pas y adhérer. Aussi, que nous ne puissions pas prendre connaissance de ceux d'une chauve-souris – pour reprendre l'exemple de Thomas Nagel – n'est peut-être lié qu'à une limite de fait, à savoir qu'il est difficile de simuler, implémenter le fonctionnement du cerveau d'une chauve-souris en train de faire de l'écholocalisation au sein de notre propre cerveau. Ce ne serait donc pas une limite logique, conceptuelle ou métaphysique.

Ensuite, il convient de ne pas faire de confusion au niveau du vocabulaire. Dans une perspective wittgensteinienne, le donné sensible peut quand même très bien avoir un rôle dans la connaissance, mais ce rôle n'est pas fondationnel car on ne peut effectivement pas

178 LEIBNIZ, Gottfried Wilhelm. *Monadologie*

formuler de propositions à son sujet. Ce n'est certes pas le donné sensible que l'on perçoit, mais c'est bien à travers lui et grâce à lui que l'on perçoit. Ce n'est certes pas le donné sensible que l'on connaît, mais c'est bien à travers lui et grâce à lui que l'on connaît. Et là nous retrouvons nos intuitions, puisqu'on dira bien que l'on rend un culte à Zeus, que l'on aperçoit un oasis au loin ou que l'on aime sa femme, et non pas, respectivement, que l'on rend un culte à une idée, que l'on aperçoit une sensation d'oasis au loin ou que l'on aime une visée intentionnelle…

Ce passage du donné sensible à la perception et à la connaissance peut être retracé en suivant un cheminement logique. Supposons qu'une personne A émette la proposition suivante : « Cette tomate est rouge », cette proposition sera comprise et jugée vraie d'une personne B, même si l'expérience qu'elle fait du rouge correspond à l'expérience du vert de la personne A. Appelons O un certain objet public tel une tomate rouge. Appelons S un certain signe telle l'expression « une tomate rouge ». Nous voudrions pouvoir justifier le fait que le signe S désigne bien l'objet O. Autrement dit, nous voudrions pouvoir justifier l'équivalence suivante : O ⇔ S.

Pour cela nous avons besoin de six prémisses que nous justifierons brièvement. Les deux premières prémisses ont trait à l'apprentissage du langage. L'acquisition du langage suppose que l'on associe un objet du monde extérieur avec un certain signe (sonore ou graphique par exemple), par présentation répétée des deux stimuli puis raffinement progressif en présentant l'objet dans des circonstances variées avec le même signe. Mais d'un point de vue individuel et subjectif, il n'est pas dit que c'est directement un objet du monde extérieur que nous associons à un certain signe (à part dans l'optique du réalisme direct). C'est cet objet, tel qu'il nous apparaît, qui est associé au signe, tel qu'il nous

apparaît. Or, on peut émettre l'hypothèse (réalisme indirect) d'après laquelle l'objet et le signe, selon qu'on les considère dans leur apparence pour nous ou dans leur être en-soi, ce n'est pas une seule et même chose[179]. De même, on peut émettre l'hypothèse selon laquelle l'apparence qu'ont l'objet et son signe pour une personne n'a rien à voir avec l'apparence qu'ils ont pour une autre personne. Nous sommes donc justifiés à introduire un certain nombre de distinctions. Ainsi, appelons respectivement OA et SA, l'objet et le signe tels qu'ils sont perçus de manière privée par la personne A. On a donc « A perçoit O » => OA et « A perçoit S » => SA. De même, appelons respectivement OB et SB, l'objet et le signe tels qu'ils sont perçus de manière privée par la personne B. On a donc « B perçoit O » => OB et « B perçoit S » => SB. Venons-en à l'établissement de nos deux premières prémisses. La personne A, lors de son apprentissage du langage, a appris à associer OA à SA. On peut donc poser : **OA ⇔ SA (1)**. La personne B, lors de son apprentissage du langage, a appris à associer OB à SB. On peut donc poser : **OB ⇔ SB (2)**[180]. Supposons à présent, comme il est assez rationnel de le supposer, que les impressions subjectives de A et de B sont majoritairement des perceptions, et non pas des illusions ou des hallucinations. Par conséquent, non seulement « A perçoit O » => OA et « B perçoit O » => OB, mais en

[179] Il suffit par exemple de penser à la variation de la taille apparente d'un objet selon qu'on le regarde de près ou de loin.

[180] La constance de l'association entre (l'apparence de) l'objet et (l'apparence de) son signe n'a d'ailleurs pas besoin d'être réelle. Il suffit qu'elle soit… apparente (cf. l'expérience de pensée de Dennett sur les goûteurs de café). Notons également que Le processus fin, au niveau psychophysiologique, la nature et l'origine précise de cette association ne sont pas clairement élucidés mais il est néanmoins difficilement contestable que les choses se passent ainsi. Le langage, à supposer même comme Noam Chomsky ou Jerry Fodor qu'il ait une composante innée et commune à toute l'humanité sous une forme rudimentaire et subconsciente qui serait à la base de l'apprentissage de toutes les langues, nécessite toutefois l'acquisition par l'expérience d'une certaine association entre le signe et l'objet qu'il désigne. Or, en l'absence préalable de la langue à apprendre pour former des définitions qui puissent permettre de connaître de nouveaux mots à partir d'anciens déjà connus, il faut bien passer, d'une manière ou une autre, par un processus, tel qu'il est décrit ici.

général il est également vrai que OA => « A perçoit O » et OB => « B perçoit O ». Donc en général, il est vrai que **« A perçoit O » ⇔ OA (3)** et **« B perçoit O » ⇔ OB (4)**. De même, non seulement il est vrai que « A perçoit S » => SA et « B perçoit S » => SB, mais en général il est également vrai que SA => « A perçoit S » et SB => « B perçoit S ». Donc en général, il est vrai que **« A perçoit S » ⇔ SA (5)** et **« B perçoit S » ⇔ SB (6)**. Nous avons à présent nos six prémisses. Par souci de clarté, nous les récapitulons ci-dessous :

(1) OA ⇔ SA

(2) OB ⇔ SB

Versant subjectif de l'apprentissage du langage

(3) « A perçoit O » ⇔ OA

(4) « B perçoit O » ⇔ OB

(5) « A perçoit S » ⇔ SA

(6) « B perçoit S » ⇔ SB

La plupart de nos impressions viennent des objets et les représentent fidèlement.

Raisonnement :

De (1) et (5) nous tirons :

(7) OA ⇔ « A perçoit S »

De (2) et (6) nous tirons :

(8) OB ⇔ « B perçoit S »

De (1) et (3) nous tirons :

(9) SA ⇔ « A perçoit O »

De (2) et (4) nous tirons :

(10) SB ⇔ « B perçoit O »

Ensuite, de (3) et (7) ou de (5) et (9) nous tirons :

(11) « A perçoit O » ⇔ « A perçoit S »

De (4) et (8) ou de (6) et (10) nous tirons :

(12) « B perçoit O » ⇔ « B perçoit S »

Enfin, puisque, par définition, une perception implique l'existence de son objet, de (11) ou de (12) nous tirons :

(13) O ⇔ S. CQFD.

La connaissance suppose donc un langage public, dont l'usage est soumis à une certaine norme externe au sujet, et doit porter sur des objets du monde extérieur et non pas sur des sense-data ineffables et indicibles. Toutes ces remarques sont correctes. Toutes ces critiques des sense-data ou d'un langage privé sont justes. Toutefois, après les avoir lu, on reste d'une certaine manière sur sa faim et insatisfait. On a l'impression d'avoir assisté à une

sorte d'exercice de style purement intellectuel dont l'effet n'est que de démolir un homme de paille.

Ces critiques ne minent ni le caractère privé des données de l'introspection ni le modèle perceptif de l'introspection indépendamment. Elles ne font que montrer qu'il est conceptuellement impossible que les états mentaux puissent être à la fois privés et à la fois objet d'une perception interne, et donc ruinent les prétentions de fonder l'édifice de la connaissance sur une base subjectiviste ou même de connaître sa propre subjectivité.

Nous ferons plusieurs remarques :

1°) Comme nous l'avons vu plus haut, à moins de souscrire à la thèse de l'Ego cartésien, nul besoin de considérer les états phénoménaux comme essentiellement privés. Ils sont privés certes, car dans les faits nous sommes les mieux placés pour expérimenter nos propres états subjectifs, mais il n'y a peut-être pas impossibilité de principe pour qu'ils soient publiquement appréhendables. Ainsi, il y a une manière moins extrémiste de comprendre les sense data, ouverte à l'intersubjectivité et notamment à la validation intersubjective. On peut comprendre le sense datum comme la perspective selon laquelle un certain objet se donne à moi, en raison de ma situation par rapport à cet objet, c'est-à-dire en raison du type de relation – appelons-la R – que j'entretiens avec lui. Cette perspective, tout sujet la connaîtrait également en observant cet objet selon le même point de vue – comprendre en entretenant avec cet objet la relation R que j'entretiens avec lui. Cette relation R, qu'un sujet donné entretient avec un objet donné et qui définit cette perspective particulière peut être décrite de manière satisfaisante avec le langage naturel de l'intersubjectivité. Illustrons cette idée par un petit exemple :

> « Je suis à la montagne avec un ami ; en fouillant le paysage à la jumelle ; je découvre, par hasard, deux chamois sur une pente herbeuse sous un gros rocher noir ; si je repose mes jumelles, ces chamois ne m'apparaissent plus que comme deux petites taches brunes se détachant mal sur un fond vert. Maintenant, je veux que mon ami puisse à son tour les observer, et pour cela je dois le guider, lui indiquer vers quoi exactement il doit pointer ses jumelles. Je vais alors lui dire quelque chose comme : " tu vois ce grand rectangle vert en face, sous le gros triangle noir ; eh bien, dans le haut à gauche, il y a deux petites taches brunes. " Et mon ami va me dire " Je ne vois rien … Ah si ! ça y est. Je vois tes deux taches. "
>
> Puis il prend ses jumelles, les oriente et observe les chamois. »[181]

Il y a donc une différence entre les chamois, en tant que tels, que l'on peut définir de façon objective sans référence à un observateur, et les mêmes chamois, tels qu'ils vont apparaître à un observateur, quel qu'il soit, du moment qu'il se place dans une certaine relation avec eux. Il n'importe pas, ici, que cette perspective soit effectivement expérimentée de telle ou telle façon dans l'intimité ineffable et indicible de l'observateur. R est parfaitement bien défini, quel que soit l'effet éminemment subjectif que cela peut bien faire d'observer un objet O sous une relation R pour un sujet S, de même qu'une table est un objet O bien défini, quel que soit l'effet éminemment subjectif que cela peut bien faire de l'observer sous une relation R1, R2, R3, etc. pour un sujet S. Et ceci est vrai, même s'il n'y a pas un tel effet éminemment subjectif.

2°) Poussons notre considération sur les perspectives plus loin. Si l'observateur est myope, daltonien, ou sujet à une maladie mentale lui empêchant de voir les objets comme des touts (prosopagnosie), la description de la relation R ne suffira pas pour voir les chamois. Mais il n'est pas encore besoin pour autant de s'échapper dans les sphères

[181] Cet exemple est emprunté à : ROSAT, Jean-Jacques. « D'un bon usage des sense-data ».

éthérées de l'ineffable… Nous pouvons définir un autre genre de perspective, toujours intersubjectif qui, si il est moins facilement connaissable par n'importe quel sujet, reste connaissable en principe.

La relation R, cette fois-ci, en plus de contenir une description de la relation qu'un sujet entretient avec un objet donné, doit aussi comporter un certain nombre de spécifications sur le sujet qui entre dans cette relation. Ces spécifications peuvent être physiques, psychologiques, cognitives, etc. mais elles sont en principe déterminables objectivement et on peut sans doute les réduire à une certaine caractérisation des relations internes entre les constituants élémentaires de l'individu, c'est-à-dire les ramener à une certaine caractérisation de sa structure fine, physique. Par conséquent, il n'est pas insensé d'imaginer (même si ce n'est pas à la portée de notre technologie…) qu'un sujet dont la structure moléculaire voire atomique serait définissable comme étant une certaine structure ST1 puisse subir une transformation, encore une fois parfaitement définissable en termes objectifs, de manière à lui faire acquérir une certaine structure ST2 qui correspondrait, par exemple, en tout ou partie, à la structure fine d'un autre sujet, de manière à lui faire voir les choses du point de vue de cet autre sujet. Là encore, il n'y a rien d'essentiellement caché, indicible ou ineffable, donc. Si on a l'impression contraire, c'est peut-être parce que ce n'est pas à notre portée de changer ainsi les individus en profondeur.

On peut cependant, dans une certaine mesure, comprendre de quoi il retourne. Tout d'abord, certaines stratégies permettent de contourner les difficultés qu'il y a à adopter une structure interne autre que celle que l'on a, afin de comprendre ce que c'est que de vivre la vie dans la peau d'une autre personne. Par exemple, on peut :

- Utiliser des accessoires, comme regarder à travers un verre rempli d'eau pour comprendre ce que c'est que la myopie, du point de vue du myope

- Utiliser son imagination, comme s'imaginer que l'on se réveille un matin avec un corps du sexe opposé pour comprendre la souffrance des transsexuels

- Imiter (au moins dans une certaine mesure), comme par exemple boire de l'alcool pour comprendre ce que c'est que d'être ivre, du point de vue de l'ivrogne

- Etc.

A vrai dire, si une telle perspective était ineffable, indicible, strictement privée et cachée à toute personne autre que le sujet qui l'expérimente, cela signifierait que les échanges d'humeur et l'empathie sont de pures illusions. Il semble pourtant que ce ne soit pas le cas et en l'absence d'arguments suffisants du contraire, il n'y a pas de raisons de supposer que ce ne sont pas des phénomènes bien réels. Le simple fait qu'une personne A puisse comprendre une personne B lorsque celle-ci lui confie ses états mentaux nécessite que ceux-ci soient communicables et aient un sens public. Si A dit à B qu'il ressent de la tristesse, de l'amertume, de l'allégresse, de l'amour, ou s'il lui dit qu'il entend mal les sons graves, qu'il trouve que le litchi a le goût de la fraise, qu'il a du mal à voir en relief, etc., B doit pouvoir comprendre pour lui ce que cela signifie. Et il n'y a rien de mystérieux à cela. C'est en associant nos propres sensations avec leurs causes et expression naturelle, puis en

associant ces dernières avec les mots désignant les sensations, que ces derniers prennent leur sens intersubjectif et nous permettent d'échanger sur nos états d'âme[182].

3°) A quel niveau se trouvent ou se trouveraient les sense data telles que les démolissent les partisans de l'argument du langage privé et leurs alliés ? Il faudrait imaginer un cas, tel que celui envisagé plus haut, dans lequel un individu, bien que se transformant de manière à devenir physiquement indiscernable d'un autre individu, aurait des *qualia* différents bien que rien ne le laisse extérieurement supposer, du seul fait qu'il resterait numériquement distinct de cet individu. Bien entendu, si un tel cas se présentait, il n'y aurait aucun moyen de le savoir. Il n'y a donc pas lieu de supposer qu'une telle chose soit possible. Plus généralement, il n'y a pas lieu de supposer que les *qualia* d'une personne autre que moi soient différents des miens si l'ensemble de son comportement me laisse supposer qu'ils sont les mêmes. On a cru, depuis Locke, qu'une inversion du spectre des couleurs pourrait exister chez un individu, sans qu'il soit possible de le détecter. Il s'agit toutefois d'une illusion métaphysique due au fait que le lien entre les sensations de couleur et le comportement est beaucoup plus subtil. Néanmoins, les couleurs nous paraissent subjectivement être ce qu'elles sont en vertu de la place qu'elles occupent dans une certaine « grammaire des couleurs », cette grammaire définissant des règles adaptées à notre action dans le monde. Il a été montré que les relations entre les couleurs ne sont pas symétriques et on pourrait s'attendre à détecter de fines différences fonctionnelles et intentionnelles chez un individu dont le spectre des couleurs serait inversé. Comme le note Jérôme Dokic [183] :

---

[182] Cf. DOKIC, Jérôme, *La perception interne et la critique du langage privé.*

[183] DOKIC, Jérôme, *La perception interne et la critique du langage privé.*

> « [L]a douleur est un cas spécial, dans la mesure où le rapport entre la stimulation nocive et la réponse naturelle est particulièrement simple et immédiat. Pour les autres sensations, la relation entre la perception et le comportement est beaucoup plus complexe et doit être conçue sur un modèle *holiste* (et non pas atomiste). Par exemple, Pears suggère que les *contours* de couleur nous livrent des informations spatiales sur notre environnement. Par conséquent, ce sont des *groupes* de sensations chromatiques qui donnent lieu à des comportements appropriés à la situation, de sorte qu'à chaque différence chromatique ne correspondra pas forcément et une fois pour toutes un comportement spécifique.
>
> On peut ainsi disposer les différents types de « sensations » sur une échelle, par exemple de la manière suivante :
>
> I. La douleur (« J'ai mal »).
>
> Les autres sensations corporelles (« Je ressens un picotement »).
>
> Les « qualités » d'orientation (« C'est à gauche/à droite/en haut/en bas, etc. »).
>
> II. Les qualités premières (« C'est une sphère, c'est grand/petit, etc. »)
>
> III. Les qualités secondes (« C'est rouge »).
>
> Plus on monte dans l'échelle, plus les relations naturelles entre la perception et l'action sont complexes et indirectes, et plus le sens des termes concernés devient « théorique ». C'est à propos des sensations qui sont à l'extrémité théorique de l'échelle qu'il est le plus facile de négliger les origines prélinguistiques, naturelles du langage. Par exemple, on croit pouvoir imaginer une inversion du spectre des couleurs qui n'aurait aucune répercussion sur le comportement du sujet. C'est oublier les liens étroits qui unissent les qualités secondes et les qualités premières, et le rapport intrinsèque de ces dernières au comportement. »

Si l'aberration d'un spectre des couleurs inversé, objectivement indétectable, n'est pas suffisamment claire on peut toutefois essayer l'exercice avec d'autres sensations, dont le lien avec les dispositions comportementales et l'intentionnalité est plus évident, immédiat voire même constitutif (le niveau I. dans l'échelle de Jérôme Dokic). Ainsi, tentons d'imaginer le cas suivant : une personne dont les sensations de plaisir et de douleur seraient inversées, mais qui ne manifesterait aucune différence comportementale, fonctionnelle,

intentionnelle. Lorsqu'elle serait soumise à un stimulus agréable, elle serait physiquement détendue et le rechercherait avidement mais dans les tréfonds de sa subjectivité elle expérimenterait en fait ce que les autres gens expérimentent lorsqu'ils sont soumis à une souffrance insoutenable. Inversement, lorsqu'on la soumettrait à des tortures effroyables, elle gémirait, ferait tout pour s'échapper, et pourtant, elle ressentirait au fond d'elle-même ce que les autres personnes expérimentent à la vue d'un beau coucher de soleil sur la mer… Cela paraît assez difficile à avaler. N'y a-t-il pas un petit problème quelque part avec la théorie lockéenne ?

La sensation de plaisir exprime naturellement une disposition à renouveler une certaine expérience, la sensation de faim, une disposition à s'alimenter, la sensation de froid, une disposition à se rapprocher d'une source de chaleur, etc. En l'absence d'argument solide selon lequel il pourrait en être autrement – et un tel argument ne pourrait exister, puisque les sensations, ainsi définies, sont reléguées dans les contrées lointaines de l'ineffable – nous sommes rationnellement conduits à considérer que tel n'est pas le cas.

Les états mentaux, les qualia, les sense data, dans la suite de cette thèse seront donc uniquement traités, sauf mention contraire, selon les sens plausibles définis en 1°) et 2°).

4°) Outre le mythe de l'intériorité – l'idée selon laquelle les qualia seraient essentiellement privés – un autre mythe doit être battu en brèche : celui du caractère perceptif de l'introspection. Alors que le dualisme cartésien a été abandonné par la plupart des chercheurs contemporains, et que le primat de la subjectivité n'est guère toléré qu'en pratique, il est remarquable de constater à quel point le modèle perceptif de l'introspection trouve encore de l'écho jusque dans les neurosciences. Ainsi, dans "Consciousness and the

Binding Problem"[184], article qui, par ailleurs, renonce à situer l'« œil interne » qui observerait les états mentaux en un point précis du cerveau, Wolf Singer propose cependant l'hypothèse selon laquelle la conscience phénoménale résulterait de la capacité des cerveaux évolués d' *« analyser leurs propres traitements cognitifs par itération et réapplication sur eux des mêmes opérations corticales qui sont utilisées pour interpréter les signaux du monde extérieur »*. Face à ce qui peut donc ressembler à une justification empirique de cet aspect du cartésianisme, une critique en bonne et due forme de l'incohérence conceptuelle de cette conception s'impose donc.

Ce modèle, rappelons-le, établit un parallèle entre l'appréhension des objets extérieurs et l'appréhension de nos représentations, dès lors élevées au rang d'objets internes ou objets perceptifs et cognitifs. Autrement dit, le fait même de percevoir le monde serait lui-même objet d'une perception, par une sorte d' « œil interne » donc, le fameux spectateur du fameux « théâtre cartésien » tant critiqué par Dennett. Cela permettrait entre autres de comprendre l'unité de la conscience. Toutefois, ce modèle perceptif de l'introspection est loin d'être valable. Tout d'abord pour des raisons conceptuelles. En effet, du moment qu'on ne saurait être aveugle de ses propres sensations, les illusionner ou les halluciner, les sensations ne sont donc pas des objets perceptifs (et *a fortiori* pas des objets cognitifs).

En effet, la perception, par nature, n'est jamais fiable à 100% et, de même que l'on peut se tromper en observant un objet extérieur, donc être victime d'illusion, ou pire d'hallucination, on pourrait s'attendre à ce qu'il soit possible de se tromper sur ses propres états mentaux, de pouvoir dire par exemple que l'on ressent une douleur et que pourtant

[184] SINGER, Wolf, "Consciousness and the Binding Problem".

cette sensation de douleur n'existe pas vraiment, qu'elle soit une hallucination ou une illusion. Toutefois, il est intuitivement difficile à admettre que l'hallucination ou l'illusion d'une sensation de douleur soit autre chose qu'une sensation de douleur, que l'on puisse distinguer, sur le plan des états mentaux, la réalité de la simple apparence. Comme le dit Kripke[185], *« à partir du moment où un certain phénomène est appréhendé exactement comme l'est la douleur, ce phénomène* est *la douleur. »*

Dès lors, n'est-ce pas un abus de langage de qualifier l'introspection de perception interne ? Et peut-on parler de connaissance introspective, dusse-t-elle porter, si on tient compte de la critique du langage privé, sur des états mentaux considérés comme publics ? Wittgenstein critique vivement le caractère incorrigible de la connaissance introspective, du moins si on entend par incorrigible quelque chose comme « toujours et nécessairement vrai » ou « qui ne peut pas être faux » et non pas « qui se situe au-delà de la distinction entre le vrai et le faux » ou « pour lequel les concepts de vrai et de faux ne s'appliquent pas ». Dans les *Recherches Philosophiques*[186], il remarque qu' *« [o]n aimerait dire ici : tout ce qui va me sembler correct sera correct. Et cela signifie seulement que nous ne pouvons pas parler de "correct" »*

Ensuite, il y a des raisons empiriques à faire valoir contre cette notion d' « œil interne », c'est qu'il ne correspond tout bonnement à rien sur le plan de la biologie du cerveau ! En effet :

> « quand on a analysé la manière dont le cerveau traite l'information, en place d'unité on ne trouve qu'un vaste ensemblage de circuits neuraux qui convergent vers… rien du tout ! Il n'est donc point d'unité dans le

185 KRIPKE, Saul Aaron. *Naming and Necessity*.

186 WITTGENSTEIN, Ludwig. Recherches Philosophiques, §258, p. 141.

> cerveau, pas de lieu particulier où se réaliserait une ultime convergence : La glande pinéale de Descartes, si elle existe bel et bien, ne remplit pas les fonctions d'intégration qu'il lui avait attribuées. De la même manière, aucune des zones de convergence existantes dans le cerveau (comme par exemple l'hippocampe, certaines parties du cortex pariétal et du cortex préfrontal) ne peut remplir seule la fonction d'intégrer l'ensemble des informations traitées. Et comment pourrait-il en être autrement ? Car si une telle zone de convergence ultime existait, qu'aurions nous appris, en réalité, sur les origines de l'unité de la conscience ? Quel homoncule percevrait les informations unifiées représentées dans ce lieu cérébral unique ? »[187]

On aurait donc pu mettre en question le modèle perceptif de l'introspection, mais cela n'allait pas de soi, après Descartes, Kant, Husserl, de considérer que l'introspection n'était pas une sorte de perception interne. En effet, que l'on considère l'introspection comme faillible ou non, il semble difficile, à première vue, de nier qu'elle fournit quand même une certaine connaissance sur soi-même. Cette connaissance n'étant pas purement analytique – pour reprendre une terminologie kantienne – si elle est néanmoins une connaissance authentique, elle devait donc être synthétique. Or, au cours du XVIIIème et XIXème siècle, la métaphysique classique et ses propositions synthétiques *a priori* avaient perdu énormément de crédibilité. C'était le temps des Lumières, puis du positivisme. Il devenait de plus en plus évident que toute connaissance synthétique devait, au moins en partie, s'appuyer sur l'expérience, donc sur la perception. La connaissance introspective, eut-elle à être considérée comme infaillible, devait donc être considérée comme une forme de perception. Cela permettait d'ailleurs d'envisager une base indubitable pour le fondement de l'empirisme anglais (Hume, Berkeley, Locke, etc.). Projet qui retrouva un écho jusqu'au début du XXème siècle, à travers, par exemple, la *Construction Logique du*

---

187 CLEEREMANS, Axel. « L'unité de la conscience ». *In :* CAZENAVE, Michel (Ed.). *De la science à la philosophie : Y a-t-il une unité de la connaissance ?*

*Monde* de Rudolf Carnap. De plus, il semblait évident que certains états mentaux au moins, telles que les douleurs diffuses, les sentiments diffus (colère, tension, fatigue, etc.), etc. n'avaient pas d'autres objets qu'eux-mêmes. Expérimenter ces états mentaux, c'était donc identique à percevoir les objets de ces états mentaux. Ces états mentaux et leur objet étant identiques, on pouvait parler d'objets mentaux. Le concept s'étendit à tous les autres états mentaux. En effet, l'argument du rêve ou du mauvais génie, issus de la tradition cartésienne, conduisait à mettre en doute l'existence d'objets réels, physiques, pour tous les états mentaux, qui devinrent ainsi leur propre objet (objet mental donc), les objets physiques correspondant à ces objets mentaux ne leur étant reliés que de manière contingente et douteuse.

Mais les états mentaux qui ne semblent avoir aucun objet réel n'en ont-ils vraiment pas ? Cet objet peut être complexe, et inclure tout l'organisme, ainsi que le système que cet organisme forme avec son environnement. De plus, s'ils étaient des objets perceptifs, leur perception ne nécessiterait-elle pas une sensation de second ordre ? On perçoit une chaise car il se créé en nous une certaine sensation. Pour percevoir cette sensation de la chaise, il semblerait donc logique de supposer une sensation de la sensation de la chaise. Mais cette sensation de second ordre devrait être aussi un objet perceptif et en appellerait ainsi à une sensation de troisième ordre, et ainsi de suite *ad infinitum*. Nous ne pouvons admettre une telle régression à l'infini et devons donc rejeter le modèle perceptif de l'introspection. Si nous le rejetons, les critiques wittgensteiniennes et apparentées s'effondrent. En effet, la sensation cessant d'être un objet perceptif (et donc cessant également d'être un objet cognitif), elle se situe en amont de là où les notions de vrai et de faux prennent sens. Ce qui veut donc dire aussi qu'elles se situent en amont du sens, autrement dit qu'elles n'ont pas

de sens, sont vides de contenu sémantique, ce qui n'est pas si grave d'ailleurs car c'est sur l'arrière-fond de ces sensations que tout le reste prend sens.

Nous préfèrerons, à ce modèle perceptif de l'introspection, le modèle pragmatique soutenu (en partie[188]) par Jérôme Dokic[189]. Ce modèle permet de comprendre comment l'introspection peut fonctionner par prise de conscience directe, non pas d'une sensation essentiellement privée, mais d'une disposition intentionnelle actualisée à travers un comportement public ou susceptible de l'être et liée de manière constitutive à cette sensation (par exemple, la recherche de nourriture est liée à la sensation de faim, le fait de regarder le soleil couchant et de dire « Quel beau coucher de soleil ! » est lié à la sensation visuelle d'un beau coucher de soleil, etc.).

Il semble peu approprié de considérer notre prise de conscience d'une attitude propositionnelle, comme, par exemple, que nous croyons que la Terre est ronde, comme étant fondée sur la conscience d'un objet, à savoir notre croyance considérée comme étant un état mental particulier. Il semble plutôt que cette prise de conscience, c'est une prise de conscience d'une certaine activité, ou d'une certaine disposition à agir, par exemple la disposition à dire « Je crois que la Terre est ronde » ou « La Terre est ronde » à la question « Quelle forme a la Terre ? »

Le modèle pragmatique de l'introspection étend tout simplement cette considération à l'ensemble des états mentaux, y compris donc les sensations.

---

188 Le modèle perceptif de l'introspection est tellement prégnant, notamment au sujet des qualia corporels, que même Jérôme Dokic ne parvient pas totalement à s'en défaire (cf. DOKIC, Jérôme, *La perception interne et la critique du langage privé*, § 5, 6, 7 et 8)

189 DOKIC, Jérôme, *La perception interne et la critique du langage privé.*

> « L'idée centrale du modèle pragmatique est la suivante. Supposons que l'on demande comment je peux distinguer par introspection un cas où je *vois* une voiture d'un cas où je l'*entends*. Une réponse traditionnelle est que j'observe « de l'intérieur » des éléments de mon paysage mental, et constate immédiatement que ceux-ci sont des sensations visuelles, par exemple, et non pas auditives. Selon le modèle pragmatique, cette réponse met la charrue avant les bœufs. En fait, je distingue un cas où je vois quelque chose d'un cas où j'entends quelque chose parce que je sais faire la différence entre *regarder* et *écouter* quelque chose. C'est ma connaissance non inférentielle des *activités* intrinsèquement liées à mes expériences visuelles et auditives qui me permet de différencier celles-ci. »

Jérôme Dokic fait remarquer que :

> « Dans la conception cartésienne, les phénomènes psychiques peuvent être désignés « de l'intérieur » indépendamment de leur expression naturelle dans le comportement. Mais précisément cette conception « intellectualise » l'apprentissage (et l'usage ultérieur) des termes de sensation. L'enfant introduit dans son langage un nouveau terme de sensation sur un fond naturel de liens complexes entre la perception et le comportement. Ces liens relèvent de ce que Pears appelle des « lignes de types » (c'est-à-dire des lignes sensori-motrices qui correspondent à des *types* de sensation plutôt qu'à des sensations particulières). Ces lignes de types sont solidement établies avant l'avènement du langage et prédéterminent l'usage et le sens du terme de sensation. »

Autrement dit, ce modèle préconise de considérer chaque type de sensation comme le terme médian d'une ligne de type reliant naturellement un certain stimulus à une certaine réponse. Il ne s'agit pas de béhaviorisme ou de fonctionnalisme simpliste, justement grâce à la présence de ce terme médian, dont l'existence n'est pas niée, mais le statut de perception interne contesté. La perception, c'est la première partie de la ligne de type, qui va du

stimulus à la sensation. La seconde partie, qui va de la sensation à la réponse, correspond à l'action. Comme le dit Pears[190] :

> « Wittgenstein ne cherche pas à éliminer ce qui se trouve au milieu de la ligne de type, mais à nous faire comprendre que c'est, non pas une autre section de la ligne, mais un point que les spectateurs peuvent approcher de façon asymptotique en partant d'une extrémité ou de l'autre. »

Selon Jérôme Dokic, ce modèle pragmatique présente plusieurs avantages :

1°) Il nous débarrasse de toute référence à des objets mentaux privés car les activités intrinsèquement liées aux sensations sont publiques.

2°) Il permet de conserver le caractère direct, non-inférentiel, de la connaissance introspective car en effet, de même que pour nos sensations, ce n'est pas sur la base d'une perception – interne ou externe – que nous savons que nous sommes en train d'agir intentionnellement. A ce sujet, Jérôme Dokic développe un argument qu'il reprend de Shoemaker[191] :

> « Quelles que soit les différences essentielles qui existent entre la perception interne et la perception externe, les deux formes de perception doivent avoir un noyau dur qui définit ce qu'est une perception *objective*. Ce noyau dur comprend la distinction existentielle entre l'acte de perception et son objet intentionnel. Pour garantir cette distinction, il semble nécessaire de préserver une certaine *indépendance* entre ses termes. En particulier, il faut admettre que l'existence de l'objet

---

190 PEARS, David. *La pensée-Wittgenstein : du Tractatus aux Recherches Philosophiques*.

191 SHOEMAKER, Sidney. *The First-Person Perspective and Other Essays*.

> intentionnel est logiquement indépendante de sa perception par le sujet. »

Or, si la connaissance du fait que nous agissons intentionnellement se faisait sur la base d'une perception, nous pourrions être aveugles à nos propres actions ou tentatives d'actions intentionnelles, et nous pourrions illusionner ou halluciner nos propres tentatives d'actions intentionnelles. Il ne semble pourtant pas que cela puisse être le cas. Comme le fait remarquer Roger Pouivet :

> « Seul celui qui connaît la fin possède le principe de son action. Autrement dit, seul celui qui agit en fonction de raisons qu'il connaît peut être dit vouloir. Et cela, au sens où il serait capable d'indiquer ses raisons au cas où on les lui demande […]. »[192]

Et :

> « L'action ne peut être dite intentionnelle que si elle est aussi décrite comme celle d'un être qui connaît la fin de son action. La décrire comme choisie revient à considérer qu'un syllogisme pratique a porté sur les moyens. »[193]

Par conséquent, on ne peut pas agir intentionnellement sans savoir qu'on agit intentionnellement. On ne peut pas plus halluciner ou avoir l'illusion qu'on tente d'agir intentionnellement. Car cela voudrait dire qu'on est en mesure d'indiquer des raisons à cette tentative illusoire ou hallucinée. Et donc, cela signifierait qu'un raisonnement pratique a eu lieu pour déterminer des moyens – dans lesquels cette tentative illusoire ou hallucinée s'inscrit « miraculeusement » – ajustés à ces fins... Mais, qu'est-ce donc, sinon une véritable

---

192 POUIVET, Roger. *Après Wittgenstein, saint Thomas*, p. 96.

193 POUIVET, Roger. *Après Wittgenstein, saint Thomas*, p. 99.

tentative, qu'une tentative si bien justifiée par celui qui l'exerce et si bien adaptée à ses fins ? Comme le dit encore Wittgenstein[194] : *« On pourrait dire […] que le mouvement volontaire est caractérisé par l'absence de l'étonnement. »*

3°) Par conséquent, les perspectives d'élaborer une théorie unifiée de la connaissance directe de nos propres états et événements mentaux seraient finalement (un peu) plus réjouissantes que prévues, du moment où l'on puisse effectivement s'assurer dans tous les cas de l'existence d'un rapport constitutif (plus ou moins direct) entre les phénomènes psychiques et le comportement et puisque l'existence de ce rapport suppose l'existence de critères publics au langage des sensations à travers les comportements et permet de faire usage de définitions déictiques publiques par désignation des comportements associés aux sensations.

## 3. Réhabilitation d'une forme de phénoménalisme

Les critiques que nous avons menées pour le moment à l'encontre du langage privé, de la perception interne, de l'incorrigibilité des états mentaux, etc. pourraient rendre vain tout phénoménalisme. On pourrait croire qu'il est désormais impossible de soutenir validement l'existence d'états qualitatifs irréductiblement subjectifs. Certes, les qualia ne peuvent recevoir une définition acceptable, qu'elle soit propositionnelle ou ostensive. Cela ne veut pas dire qu'ils ne peuvent pas être l'objet d'un certain savoir : un savoir pratique. De nombreux savoirs ne peuvent s'acquérir que par la pratique personnelle. Il serait même incongru de prétendre les apprendre dans un livre ou en observant. Ainsi, on n'apprend pas à nager en regardant des nageurs ou en lisant des livres sur la natation. Bien sûr, cela peut

[194] WITTGENSTEIN, Ludwig. Remarques sur la philosophie de la psychologie, I, § 628.

donner une idée approximative, mais il faut s'essayer soi-même à nager, procéder par essai-erreur, s'exercer, s'entraîner, pour apprendre à nager. Si apprendre la botanique n'apprend pas nécessairement à « botaniser », apprendre la natation apprend à nager. Peut-être, alors, qu'apprendre la conscience phénoménale apprend à « conscientiser phénoménalement ». Tentons l'expérience. Apprenons à « conscientiser phénoménalement » :

Considérez[195] cette image :

© Copyright Walt Disney 1928

**Phase 1 :** Posez-vous la question suivante : voyez-vous Mickey ou du moins avez-vous l'impression ou la croyance que vous le voyez ?

**Phase 2 :** Si oui, désintéressez-vous à présent de la question de savoir si cette sensation que vous voyez Mickey est une perception authentique, une illusion ou une hallucination pour vous concentrer sur la sensation visuelle elle-même, dans son

[195] Nous nous permettrons ici d'interpeller le lecteur à la deuxième personne puisque l'expérience de pensée qui suit doit être effectuée par tout un chacun pour lui-même.

immédiateté et sa spontanéité. Plus exactement, désintéressez-vous de la question de la nature du substrat métaphysique sous-jacent à cette image. Existe-t-elle toujours quand personne ne la perçoit ? Est-elle objectivement telle que je perçois qu'elle est ? Se ramène-t-elle à des bits d'informations présents dans un ordinateur géant ? Etc. Telles sont les questions auxquelles nous ne souhaitions pas que vous vous arrêtiez ici. Plongez-vous entièrement dans l'examen du phénomène lui-même pour le moment, le phénomène dont vous faites l'expérience, tel que vous l'expérimentez et en tant que vous l'expérimentez, sans spéculer sur ce que Kant appellerait sa réalité nouménale[196]. De même, désintéressez-vous également de la question de savoir si cette sensation est une propriété d'un quelconque Ego cartésien ou celle d'une activité intentionnelle enracinée dans une certaine disposition comportementale d'une personne humaine telle ou telle. Vivez-là pleinement, pour elle-même.

**Phase 3 :** A présent, posez-vous la question suivante : cette sensation ne se découpe-t-elle pas en des sensations de noir, de rouge, de jaune, de rose et de blanc, couvrant des zones de tailles variables disposées d'une certaine manière ?

Si oui, alors vous devez admettre que certaines sensations sont décomposables en sensations plus élémentaires.

**Phase 4 :** Si vous le pouvez, poursuivez cette analyse plus en avant, pour chaque constituant que vous avez isolé. Essayez en même temps de ne pas perdre de vue la sensation de départ (la vision ou l'impression visuelle de Mickey) et les différents constituants déjà dégagés (zones noires, rouges, jaunes, etc.).

---

[196] KANT, Emmanuel, *Critique de la Raison Pure*.

Il est possible que vous ne parveniez pas à garder une vision d'ensemble. Pourtant vous sentez que quelque part, les éléments que vous avez dégagés, vous en faites quand même confusément l'expérience lorsque vous vous focalisez sur la sensation globale et qu'inversement, vous faites toujours confusément l'expérience de cette sensation globale, lorsque vous vous concentrez sur les éléments qui la constituent. Est-ce une illusion ? Peu importe en fait, puisque c'est comment vous le vivez qui nous intéresse ici. En tout cas, il semble bien difficile de contester que sans aucun des éléments qui constituent votre expérience visuelle de Mickey, vous ne le verriez plus, ni n'auriez l'impression ou la croyance de le voir. Inversement, votre expérience visuelle de Mickey ne pourrait pas se transformer en, par exemple, une expérience visuelle de Donald sans qu'aucun des éléments constitutifs de votre expérience visuelle de Mickey ne change. Nous supposerons que votre mémoire n'a pas été altérée et que pour vous, une image de Donald ressemble grosso modo à ceci :

© Copyright Walt Disney 1934

A un moment, on arrive en principe à une limite : des sensations non-décomposables en sensations plus élémentaires. C'est ce qu'on appellera des qualia.

Vu comme ça, il semble difficile de nier avoir des qualia (et donc, une conscience phénoménale), et, même si nous n'avons pas donné de définition, il semble difficile de dire que nous n'avons pas acquis une certaine notion des qualia. Notion que l'on peut peut-être comparer au « nager » d'un nageur expérimenté. Matt Biondi serait sans doute désemparé si on lui demandait de définir *sa* nage. Certes, il pourrait dire, en y réfléchissant, qu'il fait tel mouvement de telle manière selon telle ou telle circonstance, et faire quelques démonstrations à un public. Peut-on dire, dès lors, qu'un public suffisamment intelligent aura compris ce qu'est *sa* nage ? Suffirait-il alors, à ce public, de s'efforcer d'imiter ces démonstrations et de suivre les consignes données pour reproduire *sa* nage ? Non, bien évidemment. De même, on peut sans doute préciser un certain nombre de conditions de possibilité de l'expérience des qualia. A elles seules, ces conditions de possibilités ne permettent pas de saisir entièrement la notion de qualia. Pour comprendre ce que sont les qualia, il faut s'y exercer. Semblablement, on ne comprend vraiment toute l'ampleur du fait de vivre un orgasme que s'il on en vit un soi-même, et l'aveugle-né ne peut comprendre de la même manière que la personne qui voit ce qu'est la lumière.

Nous nous permettrons d'apporter quelques précisions, concernant la nature des consignes que nous avons données dans notre exercice de « découverte » des qualia, afin de répondre à certaines objections. Toutes ces consignes doivent être comprises en un sens métaphysique et non pas physique.

Il ne s'agit évidemment pas d'affirmer que le fait de croire ou non en l'existence réelle l'image de Mickey et d'avoir à l'esprit tout un ensemble de significations qui lui sont reliées n'est d'aucune importance et n'a aucune incidence sur la perception que vous en avez, il ne s'agit donc pas d'un exercice contre-nature qui consisterait à modifier votre attitude naturelle vis-à-vis de cette perception qui vous est demandé. Nulle mise entre parenthèses du monde, nulle *épochè* husserlienne n'est ici exigée. Prenez cette image telle qu'elle vous apparaît, **et** telle que sa nature vous apparaît.

Nous voulons juste vous faire mettre le doigt sur une chose qui ne devrait pas changer, quelle que soit la définition de la réalité à laquelle vous souscrivez. C'est-à-dire, une chose qui sera vraie (et même nécessairement vraie), que vous soyez réaliste naïf (les choses sont telles qu'il m'apparaît qu'elles sont, et les objets colorés sont réellement colorés par exemple), ou tout au contraire idéaliste (les choses n'existent pas vraiment, seules mes pensées au sujet des choses existent vraiment), et avec tous les degrés entre ces deux positions. Sans doute y aurait-il un moyen raisonnable de savoir si la sensation visuelle que vous avez est une perception authentique, une hallucination ou une illusion : en clignant des yeux, en consultant votre ophtalmologiste, en regardant l'image sous différents angles, en la prenant en photo puis en examinant la photo, etc. Mais une telle certitude, fût-elle inébranlable, demeure *relative*. Il est toujours possible d'imaginer un moyen, aussi farfelu que l'on voudra, qui pourrait permettre d'obtenir le même résultat qu'une perception normale, sans qu'il s'agisse d'une perception normale. C'est probablement là, à notre humble avis, tout le sens du doute hyperbolique cartésien. Selon Roger Pouivet, on ne peut pas vraiment douter que le monde extérieur existe mais simplement faire semblant d'en douter (cf. les passages, déjà cités p. 163 et correspondant aux notes 108 et 109). Sans

doute, pour autant, il y a peut-être des choses pour lesquelles sembler en douter n'est pas non plus possible.

Mais il ne s'agit pas ici de nous forcer littéralement contre notre nature, a douter de l'existence du monde et donc de l'existence de l'image de Mickey, il s'agit juste de nous demander s'il serait logiquement possible, dans l'*absolu*, d'imaginer un moyen par lequel tout se passerait *comme si* le monde extérieur existât ou fût tel que nous le percevons sans que pour autant il existe véritablement ou soit véritablement tel que nous le percevons. Or, de même qu'il existe toujours un nombre indéfini de théories scientifiques qui peuvent rendre compte d'un même corpus de données empiriques, il y a tout lieu de supposer qu'il existe un nombre indéfini de théories métaphysiques qui peuvent rendre compte du corpus de toutes les données empiriques.

On peut présenter les choses autrement, et éviter ainsi toute présomption d'un scepticisme intégral. Si le monde extérieur n'existe pas vraiment et qu'il existe un certain *quelque chose* qui nous donne l'impression permanente, forte et indubitable qu'il existe, alors, en fin de compte, n'est-il pas logique de dire que le monde extérieur existe et qu'il s'identifie à ce *quelque chose* ? Si le monde extérieur n'est pas tel qu'on le perçoit et qu'il existe un certain quelque chose qui nous donne l'impression permanente, forte et indubitable qu'il est tel qu'on le perçoit, n'est-il pas logique de dire que le monde extérieur est tel qu'on le perçoit et qu'il s'identifie à ce *quelque chose* grâce auquel on le perçoit tel qu'on le perçoit ? Ainsi présenté, le doute hyperbolique n'est plus une mise en question de

l'évidence de l'existence du monde extérieur mais une interrogation légitime sur sa nature fondamentale, sur ses soubassements métaphysiques[197].

Que pourrait être ce *quelque chose* ? Diverses théories ont circulé sur sa nature. Nous rappellerons brièvement quelques classiques et mentionneront quelques développements récents. Dans la catégorie des classiques, nous rappellerons : le physicalisme métaphysique ou réalisme scientifique qui postule que la réalité fondamentale est physique[198] (à ne pas confondre avec le physicalisme méthodologique de la méthode scientifique qui est compatible en principe avec n'importe quelle option métaphysique, même le spiritualisme ou l'idéalisme), le réalisme naïf dit aussi réalisme du sens commun[199] (qui est la position qui s'accorde le mieux avec nos intuitions ordinaires sur la nature du monde), l'idéalisme

---

197 Voir CHALMERS, David John. "The Matrix as Metaphysics". L'article a été rédigé tout d'abord à l'intention de la section philosophie du site officiel du film *Matrix*

198 C'est pour beaucoup de physicalistes la position métaphysique la plus simple, minimale, qui s'impose naturellement comme métaphysique par défaut par usage du Rasoir d'Occam. Ce serait oublier que nombre de phénomènes, tels que la conscience, même si on peut ultimement leur donner une explication physique, n'entrent pas naturellement dans le cadre d'une métaphysique physicaliste et que les penser en termes physiques implique un certain effort intellectuel, le dépassement de certaines intuitions *a priori*.

199 Sous le nom technique de réalisme naïf se cache l'ontologie spontanée et de prime abord parcimonieuse selon laquelle le monde extérieur est ultimement tel qu'il nous apparaît dans la vie de tous les jours : le ciel est réellement bleu, les chaises sont réellement solides, etc. Cette position, qui est généralement le point de départ de la science, est généralement troquée en cours de route de la démarche scientifique par le physicalisme métaphysique (réalisme scientifique), plus conforme aux résultats de la science qui nous montrent que le ciel n'est pas vraiment coloré et que la solidité des chaises est apparente car de grands vides séparent ses atomes.

(monisme spiritualiste) [200], le dualisme[201] et le phénoménalisme (parfois associé à un monisme neutre ou à un panpsychisme).

Parmi les idées récentes au sujet de la nature de la réalité, signalons l'argument de la simulation[202], du philosophe Nick Bostrom, qui suggère que nous vivons peut-être dans une simulation informatique créée par une civilisation avancée (une civilisation « post-humaine » selon les termes du philosophe). Quant à Bruno Marchal, ses travaux[203]

---

200 Pour Georges Berkeley la nature ultime de la réalité est mentale (*esse est percipit*, être c'est être perçu). Seuls existent les pensées présentes dans nos esprits et que Dieu se charge de coordonner entre elles en percevant toutes choses quand bien même personne d'autre ne les percevrait, afin de les maintenir dans l'existence. Il serait erroné de voir dans l'idéalisme immatérialiste de Berkeley un scepticisme radical (ou pire un négationnisme) à l'égard de l'existence d'une réalité extérieure que nous expérimentons. Dans BERKELEY, George. *Trois dialogues entre Hylas et Philonous*, Philonous, qui défend la thèse de Berkeley, dit ceci :

> « Le bois, les pierres, le feu, l'eau, la chair, le fer, et les choses semblables que je nomme et dont je parle, sont des choses que je connais. Et je ne les aurais pas connues si je ne les percevais pas par mes sens ; et les choses perçues par les sens sont perçues immédiatement; et les choses perçues immédiatement sont des idées ; et les idées ne peuvent exister en dehors de l'intelligence : leur existence consiste donc dans la perception qu'on en a ; quand donc elles sont perçues actuellement, il n'y a pas moyen de douter de leur existence. Loin de nous tout ce scepticisme, tous ces doutes ridicules de la philosophie ! Quelle plaisanterie, qu'un philosophe mette en question l'existence des choses sensibles jusqu'à ce qu'elle lui ait été prouvée par la véracité de Dieu ; ou qu'il prétende que notre connaissance sur ce point n'atteint point à la certitude de l'intuition ou de la démonstration ! Je pourrais aussi bien douter de ma propre existence, que de l'existence de ces choses que je vois et que je touche actuellement. »

201 Pour René Descartes (DESCARTES, René, *Méditations Métaphysiques*), la nature fondamentale de la réalité est double : esprit pour l'âme humaine et matière pour le reste (corps, animaux, plantes et objets naturels ou artificiels).

202 Voir BOSTROM, Nick, "Are You Living In a Computer Simulation?" qui est consacré à la défense de cet argument. Selon ce dernier, au moins une des propositions suivantes est vraie :

(1) l'espèce humaine va vraisemblablement s'éteindre avant d'atteindre un stade post-humain (i.e. maîtrisant suffisamment la technologie pour pouvoir procéder à des simulations informatiques de leur propre histoire évolutive, ou de variantes de celles-ci) ;

(2) il est improbable qu'une civilisation post-humaine quelconque soit intéressée par la mise en route effective d'un nombre significatif de simulations informatiques de leur propre histoire évolutive, ou de variantes de celles-ci ;

(3) nous vivons presque certainement dans une simulation informatique.

203 MARCHAL, Bruno. *Calculabilité, Physique et Cognition*.

tendraient à montrer que : (1) l'hypothèse du mécanisme est probablement vraie ; (2) si l'hypothèse du mécanisme est vraie, alors la physique se réduit à la psychologie qui se réduit à la logique ; donc (3) la physique se réduit à la psychologie qui se réduit à la logique. Bref, la nature ultime du monde est logique. Enfin, nous mentionnerons une résurgence du déisme[204] à travers par exemple le thomisme analytique de philosophes comme Richard Swinburne[205].

### *C. Perspective ontologique*

Le triangle infernal de la philosophie de l'esprit représente les rapports problématiques qui opposent la perspective phénoménologique (en première personne) à la perspective physicaliste (en troisième personne) dans l'explication des comportements.

Les analyses de Wittgenstein, Ryle et autres philosophes inspirés par le béhaviorisme logique et opposés au mentalisme ainsi qu'à son corollaire : le langage de la pensée, nous ont montré que la vie mentale ne pouvait se comprendre clairement que comme une certaine disposition (plus ou moins complexe) à agir de telle ou telle façon lorsque telle ou telle condition est vérifiée.

Cette extériorisation du mental dans la sphère publique peut être poussée plus loin, et intégrer une dimension ayant été largement négligée jusqu'à lors, même par les béhavioristes : l'environnement physique et social.

---

[204] Le déisme est la croyance en Dieu appuyée sur des arguments qui se veulent rationnels et/ou empiriques, et non sur la foi et le dogme. Voir PAINE, Thomas. *The Age of Reason.*

[205] Voir par exemple : SWINBURNE, Richard. *Is there a God?*

On peut trouver plusieurs raisons à ce dédain : tout d'abord, il y a une forte intuition à vouloir considérer le cerveau, d'une manière ou d'une autre, comme étant le siège de la pensée. Peut-être que l'esprit n'est pas le cerveau, ainsi que le pensent les dualistes, mais, en tout les cas, une chose est sûre : l'esprit est dans le cerveau. Même les dualistes les plus radicaux ne peuvent nier qu'une altération du cerveau a des répercussions directes sur la vie mentale. Le physicaliste a un argument supplémentaire à faire valoir : il semble que les états mentaux puissent expliquer causalement le comportement ; or, si ceux-ci n'étaient pas identiques à un état cérébral, comment pourraient-ils en être la cause réelle ? Pourtant, cette certitude internaliste fut mise en cause, à partir de la seconde moitié des années 1970, par un nombre croissant, tant de philosophes que de scientifiques impliqués dans l'étude de l'esprit.

L'externalisme – puisque tel est le nom qu'il convient de donner à ce courant de pensée opposé à l'internalisme – commença, assez timidement, de façon purement sémantique. Putnam[206] et Burge[207], développèrent d'importantes expériences de pensée visant à démontrer que la signification des termes du langage est, *littéralement*, dans le monde extérieur : dans l'objet désigné (Putnam) ou la communauté linguistique (Burge). Il y a un petit pas en plus par rapport à Wittgenstein et à Ryle, et aucun compromis avec l'internalisme cartésien ne semble ici possible. Pour Wittgenstein, la signification des termes du langage dépendait certes de facteurs externes, mais il demeurait possible de s'en accommoder, en considérant que ceci n'empêchait en rien, sur un plan ontologique, d'identifier malgré tout ces contenus sémantiques à des états internes.

---

206 PUTNAM, Hilary. "The meaning of 'meaning'". *In:* PUTNAM, Hilary. *Mind, Language, and Reality*.

207 BURGE, Tyler, (1979), "Individualism and the Mental", *In:* FRENCH, Peter A., UEHLING, Thomas E. and WETTSTEIN, Howard K. (Eds.). *Midwest Studies in Philosophy IV: Studies in Metaphysics*.

Putnam, avec son expérience de pensée de la Terre-Jumelle, montre en effet que deux individus avec des états internes *identiques* peuvent avoir des pensées dont le contenu est *différent*.

En effet, supposons qu'il existe une planète en tout point semblable à la Terre. Nous l'appellerons Terre-Jumelle. Sur cette planète, se trouve $Oscar_2$, qui est le double ou *Doppelgänger* d'$Oscar_1$, un habitant de la Terre. On suppose que la constitution physique et les états neurophysiologiques d'$Oscar_1$ et d'$Oscar_2$ sont qualitativement identiques. Par conséquent, en vertu du principe de supervenience, selon lequel toute différence d'état psychologique entraîne une différence neurophysiologique, les deux jumeaux $Oscar_1$ et $Oscar_2$ sont donc aussi dans les mêmes états psychologiques. Or, il se trouve que sur Terre et sur Terre-Jumelle, le mot « eau » a une extension différente ; sur notre Terre, le mot renvoie à une substance dont la formule est $H_2O$, tandis que sur Terre-Jumelle il réfère à un corps dont la constitution s'exprime par XYZ. Nous sommes donc en présence d'une situation où le fait de partager les mêmes états psychologiques lors de la production du mot « eau » ne suffit pas à déterminer l'extension du mot « eau ». Comme l'extension fait partie du sens d'une expression, les états psychologiques ne suffisent pas à déterminer le sens, d'où le leitmotiv de Putnam d'après lequel « le sens n'est pas dans la tête ». Le sens est la résultante d'une élaboration sociale, où les experts ont le rôle privilégié de définir les conditions d'application d'un terme, du moins quand celles-ci sont univoques, et d'une relation indexicale, par exemple du mot « eau » à des échantillons servant de paradigme.

Un externalisme plus hardi apparu ensuite, que l'on peut appeler *externalisme actif* par opposition à l'*externalisme sémantique* de Putnam et Burge que l'on pourrait considéré comme *passif*. En effet, si Putnam et Burge identifient, ontologiquement, le contenu des états

mentaux à des éléments de l'environnement, on ne peut pas dire que cette identification apporte quelque chose concrètement, à la compréhension du comportement et de l'esprit, dans son fonctionnement. $Oscar_1$, habitant de la Terre, et $Oscar_2$, son *Doppelgänger* habitant Terre-Jumelle ont exactement le même comportement et leurs états mentaux sont phénoménologiquement identiques.

Un pas supplémentaire est franchi avec l'*externalisme actif*, encore appelé l'*extended-mind hypothesis* (l'*hypothèse de l'esprit étendu*), l'*hypothèse de la cognition distribuée*, l'*environnementalisme* (mais le terme prête à confusion avec le souci écologiste de protection de la nature…), l'*externalisme cognitif*, le *vehicle externalism* (que nous traduirons par *externalisme véhiculaire*) ou enfin – mais uniquement dans une perspective davidsonienne – la *théorie de la survenance globale*.

L'idée commune qui sous-tend ses diverses appellations est la suivante : l'environnement d'un individu fait partie intégrante de son système cognitif. Autrement dit, ce ne sont pas seulement les contenus des états mentaux qui sont dans le monde extérieur, ce sont aussi les états mentaux eux-mêmes – tout au moins certains d'entre eux et en partie.

Ainsi, si une personne et son *Doppelgänger* qualitativement identique se trouvent dans un environnement physique et social différent, alors ils seront cognitivement, comportementalement et peut-être aussi phénoménologiquement différents (l'hypothèse de l'esprit étendu peut ne pas concerner les aspects phénoménaux de la conscience ; même si nous verrons plus loin que supposer des qualia externes n'est pas sans pertinence). Une question vient immédiatement à l'esprit. Le cerveau n'est pas un système fermé. Même l'internaliste le plus forcené ne remettra pas en question cette donnée factuelle. Par

conséquent, plongés dans deux milieux différents, deux individus qualitativement identiques ne tarderont pas à différer. N'est-ce donc pas plutôt *en vertu* des modifications internes que le monde externe va engendrer que deux « jumeaux » microphysiques vont en venir à différer mentalement et comportementalement ? C'est sans doute vrai en partie. L'externaliste ne peut le nier. Mais la position de l'externaliste n'est pas tant de nier la pertinence de processus internes dans l'explication du mental et du comportement que de l'étendre à tous les processus externes qui sont couplés avec ces processus internes. Le comportement, la cognition et peut-être la phénoménologie d'un sujet ne seront pas les mêmes selon que seuls les processus internes auront lieu, indépendamment de tout environnement (supposons qu'ils apparaissent spontanément ou sous l'action d'un malin génie cartésien), ou selon qu'ils seront réalisés en couplage avec des processus externes d'un environnement.

Si cette hypothèse est vraie – et nous montrerons plus loin qu'elle est supportée par des arguments très forts, en dépit du caractère déroutant qu'elle peut présenter de prime abord – elle a des conséquences importantes dont une – qui n'est pas des moindres – concerne directement notre conception de la personne.

En effet, tant que l'on considère chaque esprit comme un système clos, une monade « sans porte ni fenêtre », alors on ne peut distinguer deux personnes que par des critères intrinsèques. Or il semble que deux personnes, à partir du moment où elles occupent un espace différent, sont numériquement distinctes, quand bien même elles seraient intrinsèquement indiscernables l'une de l'autre. Adopter une attitude internaliste revient à contester le bon sens de cette intuition, et à considérer qu'il ne suffit pas que deux personnes occupent des espaces distincts pour être numériquement différentes. Il faut, en

outre, considérer d'autres facteurs, lesquels ne peuvent être qu'intrinsèques (matière, âme, etc.). Et puisque ces facteurs ne feront pas de différence observable, cela revient à fonder la distinction numérique entre personnes sur des considérations métaphysiques. Or un passage si précoce dans le domaine métaphysique n'est pas requis si on admet l'externalisme actif, car chaque personne, en plus d'être caractérisée par des qualités intrinsèques, est également caractérisée, à un moment donné, par une certain rapport à l'environnement, par une certaine localisation spatiale qu'elle est la seule a avoir.

D'autres conséquences de l'externalisme actif sont épistémologiques. Elles concernent sa variante phénoménologique, c'est-à-dire l'idée selon laquelle il existe aussi, *littéralement*, des qualia étendus. Comme nous l'avons vu plus haut, selon cette hypothèse, un système isolé n'a pas la même phénoménologie que le même système – qualitativement identique – plongé dans un système plus vaste avec lequel il interagit. Par conséquent, selon que la perception d'un objet par un sujet est véridique (perception réelle) ou non-véridique (illusion ou hallucination), il faudrait s'attendre, contrairement à ce qu'une tradition sceptique tenace maintient comme une évidence depuis Pyrrhon, à ce que la phénoménologie diffère. Il deviendrait donc par conséquent possible, en ne faisant appel qu'aux ressources de l'expérience immédiate, de distinguer entre une perception réelle et une hallucination ou une illusion. Bien sûr, on peut toujours, en n'étant pas exercé, ou attentif, se tromper sur la nature épistémique de notre expérience, mais il n'y aurait pas impossibilité de principe à ce que l'on puisse la découvrir. Ainsi, contrairement à un certain préjugé sceptique, un témoignage sincère, par exemple d'enlèvement par des extraterrestres, devrait être considéré de prime abord comme une perception authentique, à moins qu'il y

ait des raisons indépendantes de le considérer comme une hallucination[208]. Il y a là quelque chose d'assez comparable à un renversement copernicien !

Pour finir, une autre conséquence majeure de l'externalisme actif, du moins si on l'étend aux aspects phénoménaux de la conscience, pourrait être une sorte de panpsychisme voire de panthéisme puisque l'on considère qu'une partie au moins des états mentaux voire des qualia sont, *littéralement*, au moins en partie dans le monde exterieur. De là à supposer

---

[208] Cette critique du scepticisme radical et pour ainsi dire ontologique du pyrrhonisme ne saurait valoir comme critique générale du scepticisme méthodologique propre à la démarche scientifique. En effet, on peut avancer des raisons scientifiques au doute tenant à un témoignage d'enlèvement par des extraterrestres. Par exemple, on peut faire valoir le fait que : l'hypothèse extraterrestre est difficilement testable (il faudrait pouvoir capturer un extraterrestre et sa soucoupe, ou bien qu'ils fassent un débarquement ostensible et officiel, etc.), qu'elle est peu économique ontologiquement (le rasoir d'Occam conduit à privilégier des hypothèses plus en accord avec nos connaissances actuelles s'il en existe), qu'elle est peu plausible au regard de nos connaissances physiques actuelles (impossibilité physique de dépasser la vitesse de la lumière qui rend peu plausible des voyages interstellaires associé au constat que la Terre est la seule planète du système solaire a disposer des conditions propices pour qu'une vie intelligente se soit développée, etc.), qu'elle pose plus de questions qu'elle n'apporte de réponses (quel est le but de ces extraterrestres ? comment s'y prennent-ils ? d'où viennent-ils ? etc.). On sait, de plus, que le mensonge reste possible et que le témoignage humain est faillible. Enfin, la possibilité d'une illusion ou d'une hallucination n'est pas exclue, du moment que cette possibilité s'appuie sur des bases scientifiques et non sur des raisons *a priori*. Un bon sceptique, actuellement, devrait donc, si l'externalisme actif est vrai et a les conséquences épistémiques que nous avons mentionnées, juste ne pas rejeter *a priori* le témoignage humain, et accepter *a priori* que celui-ci puisse être véridique, fût-il extraordinaire. C'est ce type de scepticisme non-dogmatique qui est celui qu'applique normalement les scientifiques dans leur recherche. C'est également celui que revendiquent les divers organismes et associations sceptiques enquêtant sur le domaine du « paranormal », tel le CSICOP (*Committee for the Scientific Investigation of Claims of the Paranormal*) aux Etats-Unis ou l'Observatoire Zététique en France. Il reste que ce scepticisme peut bien entendu faire l'objet d'une critique. Mais cette critique ne saurait légitimement s'appuyer sur cette conséquence épistémique supposée de l'externalisme actif. Paradoxalement, même, on notera que les sceptiques méthodologiques actuels adoptent souvent des thèses externalistes, sur un plan épistémique, sémantique ou ontologique, celles-ci prenant à contre-pied le cartésianisme, le subjectivisme, le relativisme cognitif, etc. Paradoxalement toujours, c'est vers le scepticisme métaphysique à la Pyrrhon que certains partisans de thèses opposées au scepticisme méthodologique tentent parfois de se refaire une crédibilité, oubliant que le scepticisme méthodologique n'implique pas le pyrrhonisme.

que ces états mentaux et qualia externes persistent, même en absence de tout sujet, il n'y a qu'un pas, qui est franchi par certains penseurs de la théorie Gaïa, selon laquelle la Terre serait comparable, par ses nombreuses et complexes boucles de régulation, à un organisme vivant[209]. Si une perception sans objet diffère phénoménologiquement d'une perception avec objet, il est loisible d'imaginer la possibilité d'une perception cette fois-ci *avec objet mais sans sujet*, avec des propriétés phénoménales spécifiques. Une forme de survie impersonnelle de la conscience pourrait donc être envisageable, par fusion de l'expérience individuelle (perception avec ou sans objet) dans l'expérience universelle (perception sans sujet).

[209] La théorie Gaïa (du nom de la déesse grecque de la Terre) a été initialement proposée par James Lovelock en 1969 comme une hypothèse (LOVELOCK, James Ephraim. *Gaia: A New Look at Life on Earth*). Elle a été ensuite reprise par Lynn Margulis en 1979 (MARGULIS Lynn and SAGAN, Dorion. *Microcosmos: Four Billions Years of Evolution from Our Microbial Ancestors*). A l'origine, la théorie Gaïa se borne à constater l'existence de processus d'homéostasie à échelle planétaire, ce qui est généralement plus ou moins admis par la communauté scientifique. Ce n'est que plus tard qu'apparaissent des dérives « new age » de la « mouvance Gaïa » faisant de la Terre un organisme vivant au même titre qu'une vache, voire lui attribuant de l'intentionnalité ! Mais comme le note Richard Dawkins (DAWKINS Richard, *Le Gène Egoïste*), qui confirme par ailleurs l'existence de phénomènes d'autorégulation à l'échelle planétaire, la condition essentielle nécessaire à définir un être vivant et son évolution, qui est l'opposition permanente à un milieu extérieur - proies et prédateurs -, seule susceptible de le faire évoluer au fil du temps par le mécanisme bien connu de l'évolution naturelle, n'existe pas pour la Terre, qui n'a ni milieu, ni proies, ni prédateurs. Dawkins affirme donc parfaitement abusive le fait de la faire assimiler à un organisme vivant au seul prétexte qu'elle possède une dynamique comme en a aussi n'importe quel système animé (un incendie de forêt, un ouragan, une avalanche...) et que les autorégulations concernées n'ont pas le moindre rapport avec des manifestations de vie. Il y a donc plusieurs niveaux, dans la théorie Gaïa. Certains sont tout à fait acceptables, dont la conception de James Lovelock et de Lynn Margulis. D'autres, comme ceux de la « mouvance Gaïa » faisant de la Terre un être vivant recherchant son équilibre et manipulant consciemment le climat, sont purement spéculatifs.

## 1. L'esprit étendu

Pour le moment, bien entendu, nous n'avons exploré que quelques conséquences de l'externalisme actif et, en particulier, de ce que l'on pourrait appeler l'externalisme phénoménal, mais nous n'avons pas démontré la pertinence de telles idées. A la limite, le caractère parfois incongru de ces conséquences militerait plutôt contre la véracité de l'externalisme actif et, *a fortiori*, de l'externalisme phénoménal. Toutefois, il y a de bonnes raisons qui plaident en leur faveur. Nous en examinerons quelques-unes ci-après. Ce qui est particulièrement intéressant, c'est que les arguments à l'appui de l'externalisme actif ne sont pas uniquement des arguments purement philosophiques. Certains arguments sont issus de considérations empiriques et sont le fait de scientifiques de formation et de profession.

Un article majeur et précurseur de l'externalisme actif est celui de Andy Clark et David Chalmers, *The Extended Mind*[210]. Dans cet article, les auteurs proposent de considérer les trois situations suivantes, où un être humain est mis dans une situation où il doit résoudre un problème :

1) Une personne est assise devant un écran d'ordinateur qui affiche des images bidimensionnelles de formes variées. Certaines représentent des éléments solides et d'autres représentent des cavités. Pour chaque élément solide affiché, on affiche également une cavité, et on demande à la personne d'évaluer la potentialité d'emboîtement de l'élément dans la cavité. Pour ce faire, la personne doit, mentalement, retourner chaque élément et l'aligner à la cavité.

[210] CLARK, Andy and CHALMERS, David John, "The Extended Mind".

2) Une personne est assise devant un écran d'ordinateur similaire, mais cette fois peut choisir, ou bien de faire tourner physiquement les images sur l'écran, en pressant un bouton de rotation, ou bien de les faire tourner mentalement comme avant. Nous pouvons aussi supposer, assez vraisemblablement, que la rotation physique va permettre un accroît de vitesse avantageux.

3) Quelque part dans un futur cyberpunk, une personne est assise devant un écran d'ordinateur similaire. Cette personne, cependant, a bénéficié d'un implant neural qui peut réaliser la rotation aussi vite que l'ordinateur dans l'exemple précédent. La personne doit encore choisir quelle ressource interne utiliser (l'implant ou la bonne vieille rotation mentale), puisque chaque ressource effectue différentes demandes à l'attention et autres processus cérébraux concurrents.

Chalmers et Clark prétendent ensuite que ces trois cas sont similaires. En effet, le cas (3) semble au même niveau que le cas (1), quant au cas (2), il présente le même type de computation que le cas (3), réalisé par dispositif externe ou interne respectivement. Si la rotation, dans le cas (3) est cognitive, de quel droit devrions-nous considérer le cas (2) comme fondamentalement différent ?

L'internaliste pourrait rétorquer que, même si instrumentalement, les traitements effectués par l'ordinateur dans le cas (2) ou l'implant dans le cas (3) sont cognitifs, la cognition réelle s'effectue dans le cerveau, autrement dit, que c'est uniquement en vertu des

effets sur le cerveau qui ont pour conséquence des effets cognitifs réels, que l'ordinateur et l'implant pourraient être vus comme des extension de notre système cognitif.

De plus, comment considérer comme partie intégrante d'un système cognitif, un dispositif qui ne serait pas constamment relié à ce système cognitif, qui en serait en quelque sorte désolidarisé et pourrait être couplé à un autre système cognitif ? Or dans le cas (2), l'ordinateur n'est évidemment pas constamment couplé à la personne qui l'utilise, et peut très bien servir à une autre personne.

Face à cette critique, Chalmers et Clark nous proposent une autre expérience de pensée où une telle interprétation ne semble guère tenir. On imagine le cas d'Otto, un patient atteint de la maladie d'Alzheimer, et qui, du fait de troubles sévères de la mémoire, est obligé de noter constamment dans un carnet toute information importante pour lui. Chalmers et Clark montrent ensuite que le rapport qu'à Otto avec son carnet est fonctionnellement équivalent à celui qu'à Inga, une personne normale, avec sa mémoire. En effet, supposons qu'Inga et Otto entendent dire qu'il y a une exposition au Musée d'Art Moderne, et souhaitent s'y rendre. Inga réfléchit un moment et se rappelle que le musée est à la 53ème rue, puis elle se met à marcher vers la 53ème rue et va au musée. De son côté, Otto consulte son carnet, dans lequel est noté que le musée se trouve à la 53ème rue, puis il se met à marcher vers la 53ème rue et va au musée. Il semble clair que Inga a la croyance que le musée est à la 53ème rue avant de consulter sa mémoire. De même, on doit donc dire qu'Otto à la croyance que le musée est à la 53ème rue, avant de consulter son carnet.

L'alternative serait de dire que Otto n'a pas de croyance au sujet de la localisation du musée jusqu'à ce qu'il consulte son carnet. Au mieux, il croit que le musée se trouve à

l'adresse indiquée dans le carnet. Mais si nous côtoyons Otto un bon moment, nous constaterons à quel point une telle manière de s'exprimer n'est pas naturelle. Otto est constamment en train d'utiliser son carnet tout naturellement. Il est central dans ses actions dans toutes sortes de contextes, au même titre qu'une mémoire ordinaire est centrale au sein d'une vie ordinaire. Dire que ses croyances disparaissent quand son carnet est fermé semble aussi incongru que de dire que les croyances d'Inga disparaissent aussi longtemps qu'elle n'a pas conscience de celles-ci.

## 2. Psychologie in vitro

L'argumentation de Clark et Chalmers est convaincante, mais une position internaliste reste possible. De plus, les auteurs abordent essentiellement la conscience sous un angle cognitif, mais sont silencieux quant à l'extension de l'hypothèse de l'esprit étendu aux aspects phénoménaux de la conscience. Nous allons voir cependant que l'internalisme est intenable, à moins d'admettre qu'il soit littéralement possible de faire de la psychologie *in vitro*[211]. Par la même occasion, nous monterons que la conscience phénoménale est étendue également, qu'il y a donc lieu de supposer des *qualia* étendus, en partie externes.

Supposons un internaliste extrême, pour qui toute information venant à la conscience passe nécessairement par un endroit bien précis du cerveau, mettons la glande pinéale. Une expérience de douleur à la hanche ou de vision d'un coucher de Soleil ou d'amour pour la patrie correspond donc nécessairement à un certain type d'impulsion

[211] Cette discussion s'inspire partiellement de LLOYD, Peter B. "Is the Mind Physical? Dissecting Conscious Brain Tissue". Toutefois, nous ne partageons pas les vues dualistes de l'auteur, et estimons que son argumentation ne vaut que pour les formes internalistes de physicalisme, et non pour ses formes externalistes, de plus en plus en faveur.

électrique qui va aller activer ces quelques neurones. Si une telle conception est correcte, alors il faut en tirer la conséquence suivante, qui pour inattendue n'en demeure pas moins parfaitement logique : si nous prélevons la glande pinéale d'une personne, tuons la personne et maintenons en vie artificiellement juste cette glande pinéale, alors nous avons sauvegardé l'essentiel de l'individu ! Mieux, si nous envoyons dans cette glande pinéale un flux électrique correspondant à une expérience de douleur à la hanche, de vision d'un coucher de soleil ou d'amour pour la patrie, nous sommes justifiés à dire que ce petit bout de chair fera les expériences correspondantes ! Une autre version de l'internalisme pourrait être celle-ci : toute information venant à la conscience ne passe pas nécessairement par un point précis du cerveau. Toutefois, l'expérience de la douleur à la hanche va correspondre à un endroit spécifique, l'expérience de la douleur au pied à un autre, l'expérience du rouge dans la partie supérieure droite de l'œil gauche à une autre partie, l'expérience du goût du café à une autre, et ainsi de suite. Dans cette version, le champ de la conscience est décomposé en éléments simples et relativement autonomes correspondant à des petites parties localisés et relativement autonomes du cerveau. Selon ce modèle, si un être humain souffre du pied droit, cela va correspondre à, mettons, une activation de la zone Z1, si il souffre du pied gauche, cela va correspondre à, mettons, une activation de la zone Z2, et s'il souffre des deux pieds, cela va correspondre à une activation des zones Z1 et Z2. Cette version de l'internalisme n'est toutefois pas meilleure que la précédente. Le premier problème est qu'elle rend mystérieux le caractère unifié de la conscience. Pourquoi l'activation de Z1 et de Z2 produit-elle une conscience de Z1 et Z2 et non pas une conscience de Z1 et une conscience de Z2 disjointes ? Ensuite, une psychologie *in vitro* est toujours possible. Prélevons la zone Z1 chez un sujet, maintenons-la en vie dans une éprouvette et activons-la. En toute logique, on a alors créé, dans cette éprouvette, une expérience de douleur au pied droit !

Si l'internaliste ne veut pas assumer ces conséquences logiques, il est donc forcé d'admettre que l'expérience consciente est générée par l'activité synchronisée entre les diverses parties du cerveau. La conscience n'est pas plus en un endroit qu'à un autre qu'une université n'est plus dans un de ses bâtiments que dans un autre[212]. En d'autres termes, la conscience est une fonction globale d'états cérébraux locaux. La douleur au pied droit est une valeur particulière de cette fonction, la douleur au pied gauche en est une autre, etc. Avec ce modèle, on voit que l'isolation dans une éprouvette de la zone Z1 ne permettra jamais de produire un mal de pied *in vitro*, même si l'activation de la zone Z1 peut être un paramètre essentiel à la survenue de cette expérience, à l'intérieur de la dynamique cérébrale.

Fort bien, mais en faisant cette concession, l'internaliste sape ses propres fondements. En effet, à partir du moment où cette idée selon laquelle la conscience est une fonction globale d'états cérébraux locaux est acceptée, la prémisse essentielle de l'externalisme est acceptée. Il est montré cliniquement qu'il est possible de survivre à l'ablation ou la destruction d'un hémisphère cérébral. L'hémisphère restant semble, de plus, présenter une vie mentale. Certes, cette vie mentale est drastiquement diminuée, mais il n'est nul besoin d'en revenir à un réductionnisme localisationniste appuyant une psychologie in vitro fantaisiste de tissus cérébraux pour comprendre comment une vie mentale peut exister au niveau d'un seul hémisphère. En effet, l'hémisphère cérébral est une entité suffisamment complexe pour qu'un « calcul » sur les états cérébraux locaux restant permette l'émergence d'une conscience amoindrie mais bien réelle. De plus, on sait également que les patients *split-brain*, ces personnes dont les commissures inter-hémisphériques ont été sectionnées afin de soulager certaines formes d'épilepsie,

[212] L'exemple est repris de celui de RYLE, Gilbert. *The Concept of Mind*, « Descartes' Myth », p. 11-24.

présentent, dans certaines conditions expérimentales, ce qui apparaît être un dédoublement de la conscience, l'hémisphère droit ignorant ce que fait le gauche, et réciproquement. Considérons maintenant les 4 situations suivantes :

1) Un hémisphère gauche seul dans un crâne

2) Le même hémisphère gauche relié à un hémisphère droit.

3) Une personne seule chez elle.

4) La même personne en train de discuter avec un ami.

Fonctionnellement, les cas 1 et 2 sont équivalents aux cas 3 et 4 respectivement. Autrement dit, le cas 4 est au cas 3 ce que le cas 2 est au cas 1. Un raisonnement par analogie est donc possible. Récapitulons :

1. Niveau des hémisphères cérébraux :

   1.1. L'hémisphère gauche seul a une vie mentale.

   1.2. L'hémisphère gauche en interaction avec un hémisphère droit a une vie mentale plus riche.

2. Niveau des personnes :

   2.1. La personne seule a une vie mentale.

2.2. Une conséquence semble inévitable : la personne en train de discuter avec un ami a une vie mentale plus riche.

Le constat peut sembler trivial. Il ne l'est que partiellement. Voyons pourquoi :

Si l'hémisphère gauche en interaction avec un hémisphère droit a une vie mentale plus riche, ce n'est pas seulement parce que l'hémisphère droit va modifier instrumentalement la conscience hémisphérique gauche en modifiant les états cérébraux locaux de l'hémisphère gauche en fonction desquels émergerait cette conscience hémisphérique gauche. C'est parce que la conscience hémisphérique gauche a, en quelque sorte, fusionné avec la conscience hémisphérique droite en une conscience inter-hémisphérique, une conscience cérébrale globale. Au lieu d'être une fonction globale d'états cérébraux locaux situés dans un seul hémisphère, elle devient une fonction globale d'états cérébraux locaux situés dans deux hémisphères, ce qui est tout à fait différent. Ainsi, non seulement cette conscience est partagée entre les deux hémisphères et n'est pas localisée dans un seul, mais encore, sans un des deux hémisphères, l'hémisphère restant ne pourrait, à lui seul, faire émerger cette conscience, même pas par hasard, ou en recevant d'un Malin Génie (par exemple) les *inputs* adéquats. Les activités de chaque hémisphère sont constitutives de cette conscience cérébrale.

Cette idée est maintenant couramment admise par de nombreux neurobiologistes, la recherche de ce que Daniel Dennett appelle « le théâtre cartésien » s'étant soldée par un cuisant échec. La conscience doit donc, au moins, être conceptualisée comme un phénomène global, impliquant des régions dispersées du cerveau reliées dans une dynamique d'ensemble. Mais si on admet ceci, alors il semble que nous soyons logiquement

contraints d'admettre aussi cela : c'est que si la personne en train de discuter avec un ami a une vie mentale plus riche, ce n'est pas non plus uniquement parce que cette discussion modifie son cerveau, et donc sa conscience cérébrale individuelle, c'est aussi parce que cette conscience fusionne littéralement, et au moins partiellement, avec celle de son interlocuteur, pour former une seule conscience plus vaste. Le couplage action-perception qui s'instaure entre cette personne et son ami permet de connecter fonctionnellement leur cerveau, et le système dynamique ainsi constitué fait émerger une conscience interpersonnelle globale qui est fonction des états cérébraux locaux des deux protagonistes. Nous avons donc là une claire et puissante démonstration de *la thèse de l'esprit étendu*, qui ne se base que sur des prémisses scientifiques et l'incongruité de la psychologie *in vitro* qu'implique le *matérialisme cartésien* et ses variantes. On remarquera au passage comment certaines expressions populaires prennent d'un seul coup l'allure d'intuitions géniales. Ainsi, c'est peut-être bien au sens littéral du terme que deux amoureux en parfaite communion et harmonie ne forme plus *qu'un seul être*, leur conscience individuelle fusionnant en une conscience interindividuelle dont les états sont « calculés » par une fonction prenant en compte leurs états cérébraux respectifs. Dans le même ordre d'idée, lorsque l'on dit qu'après notre mort, on survie dans la mémoire de nos proches, il faut peut-être comprendre cela de manière littérale. Nous avons vu qu'un individu, quoique diminué mentalement, pouvait survivre à la destruction d'un de ses hémisphères cérébraux. La conscience interindividuelle que cet individu pourrait former avec ses congénères, notamment ceux avec lesquelles il forme les relations les plus harmonieuses (parents, amis très proches, conjoint, etc.) pourrait donc, quoique de manière affaiblie, très bien survivre à sa disparition. Si on considère froidement l'amour comme la création d'une fonction-conscience interindividuelle d'individus qui soit harmonieuse, alors on pourrait même comprendre pourquoi tous les grands mystiques et maîtres spirituels ont toujours insisté

sur l'importance de l'amour, et y ont vu une voie sinon LA voie de salut pour l'homme, par laquelle il peut aspirer à l'immortalité. On peut également réinterpréter, à la lumière de l'externalisme actif, ce qu'il convient d'appeler *la communication avec l'au-delà*. L'externalisme actif expliquerait ainsi pourquoi l'amour des participants à une séance pour le disparu semble être un préalable indispensable pour de bonnes communications, et il permettrait de réconcilier sceptiques et tenants du spiritisme. Les sceptiques aurait raison de penser que ce qui est communiqué, lors de séances de spiritisme, se trouve intégralement dans le subconscient des participants, mais les tenants auraient également raison en pensant communiquer ainsi avec leurs chers disparus, car ceux-ci, en quelque sorte, habiteraient justement le subconscient des participants !

### 3. De l'illusion d'habiter son corps

Il reste peut-être à répondre à quelques objections. Pourquoi, si l'*externalisme actif* est vrai, a-t-on l'impression si tenace qu'il est faux ? Autrement dit, pourquoi a-t-on la vive impression d'habiter notre cerveau ? Pour le comprendre, il n'est peut-être pas inutile de s'intéresser à ce qu'en dit Dennett. Pour ce philosophe, ce que l'on appelle le « Soi » et que l'on a l'habitude de localiser quelque part entre nos deux oreilles, n'est rien de plus qu'une sorte d'illusion créée par la disposition particulière de nos récepteurs sensoriels sur notre corps. Pour montrer cela, il raconte, dans une nouvelle de science-fiction le mettant en scène[213], une expérience imaginaire où son cerveau, soigneusement déposé dans un dispositif le maintenant en vie, est relié par ondes électro-magnétiques à un robot humanoïde doté de récepteurs sensoriels et envoyé en mission dans un endroit dangereux.

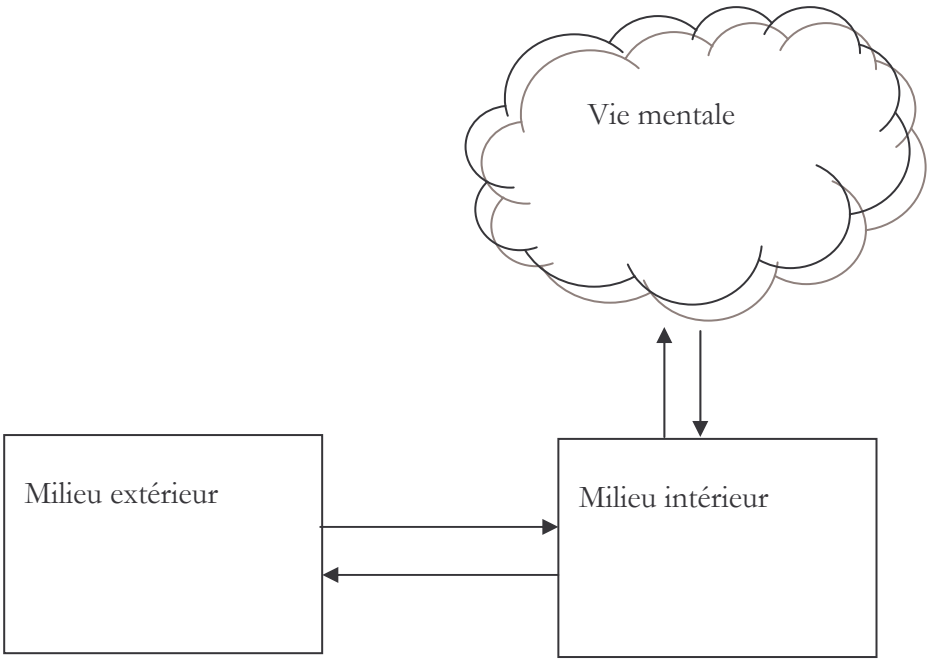
---

213 DENNETT, Daniel Clement. « Où suis-je ? » *In :* HOFSTADTER, Douglas et DENNETT, Daniel Clement. *Vues de l'esprit : fantaisies et réflexions sur l'être et l'âme.*

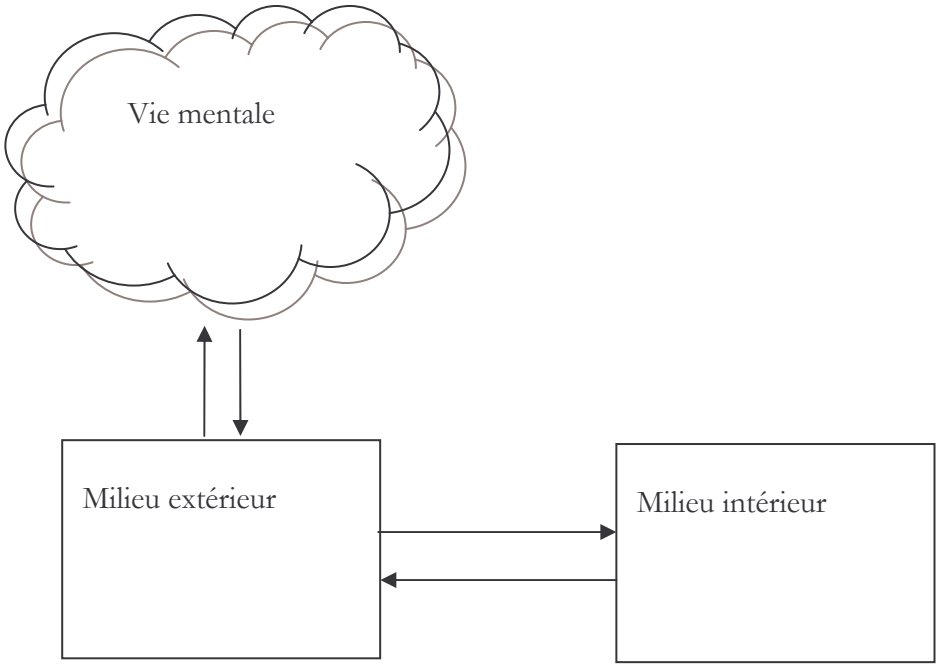
Le récit montre bien que, les connexions établies, Dennett a l'impression saillante d'être, non pas là où se trouve son cerveau, mais là où se trouve le robot, lequel pouvant très bien se trouver à plusieurs milliers de kilomètres de là où est entreposé son cerveau. L'expérience est fictive bien entendu, mais si on suit son cheminement logique en se mettant à la place du narrateur, on comprend très bien que cela se déroulerait ainsi, si l'expérience pouvait techniquement être réalisée. On aurait l'impression d'être dans le robot, et non pas là où serait notre cerveau.

La réponse à cette première objection suggère paradoxalement un autre argument en faveur de l'externalisme actif. Retournons tout d'abord à la conception internaliste. Selon celle-ci, lors d'une perception, des modifications dans l'environnement vont entraîner, par le biais des récepteurs sensoriels, des modifications dans le cerveau et ses modifications cérébrales seront identiques à des modifications mentales ou, du moins, des modifications mentales surviendront sur ces modifications cérébrales. Toujours selon l'internalisme, lors d'une action, on aura d'abord des modifications psycho-cérébrales qui vont entraîner des modifications au niveau musculaire, ce qui va entraîner un mouvement qui va modifier l'environnement.

Nous pouvons représenter ces processus sur un schéma.




Cependant, outre le fait que nous avons l'impression d'habiter dans notre tête, pour les raisons illustrées par Dennett, il semble qu'il n'y a pas plus de raison de soutenir cette représentation, plutôt que celle-ci :




En effet, milieu intérieur et milieu extérieur sont symétriques dans le rôle qu'ils jouent dans la perception et l'action. Pour bien nous en rendre compte, considérons ces quelques exemples :

On pense que le milieu intérieur est directement lié à la conscience car si on touche au cerveau, on modifie la conscience ; fort bien, mais on pourrait rétorquer qu'en touchant à l'environnement, on modifie aussi la conscience, parfois sans que celle-ci se soit d'ailleurs aperçue de quelque chose. Un mauvais cadre de travail a une influence directe sur la motivation et la productivité, mais il n'est pas dit que l'on a forcément conscience du fait que la cause de notre manque de motivation est située dans l'environnement. Il est même possible qu'on lui assigne une autre cause, interne cette fois-ci, et que l'on se culpabilise alors que l'on n'a rien à se reprocher vraiment (à part peut-être de n'avoir pas ranger son

bureau pour le rendre convivial…). Alors pourquoi placer le siège de la vie mentale dans le cerveau ?

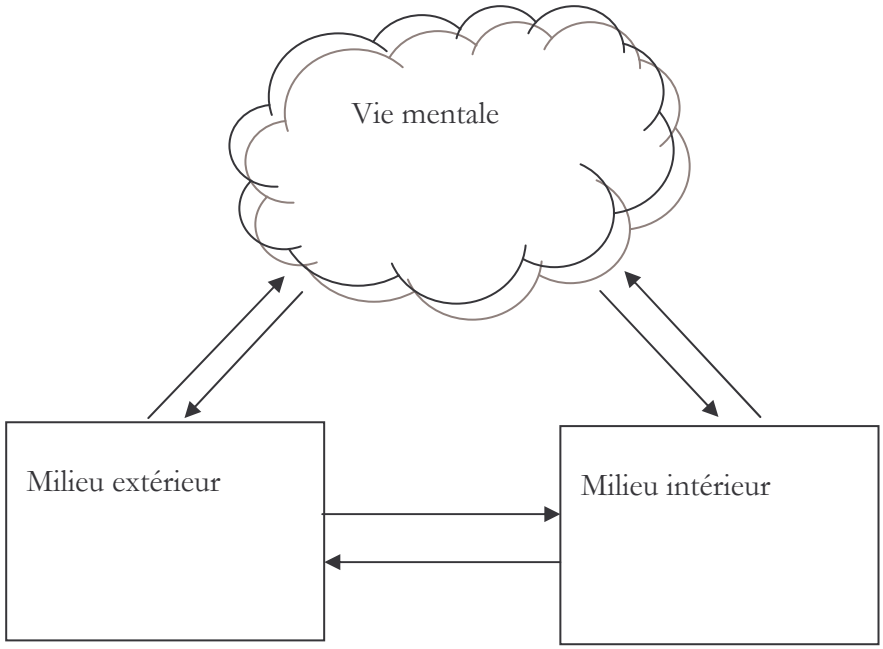
On pense que le milieu extérieur n'est pas directement lié à la conscience car c'est l'environnement que la conscience perçoit et sur lequel elle agit. Mais si, à environnement identique, on modifie quelque chose dans le cerveau, la conscience peut s'en rendre compte. Imaginons par exemple que vous êtes en face d'un mur blanc. Pendant que vous le regardez, vous savez qu'un neurochirurgien manipule votre cerveau. Soudain, vous avez l'impression que le mur est noir. Ce que vous percevez, cependant, n'est pas un changement effectif du mur puisque celui-ci est toujours blanc (objectivement) et vous le savez, puisque vous n'avez pas pour habitude de voir des murs changer instantanément de couleur, surtout quand personne n'est là pour les repeindre ! Ce que vous percevez, c'est plutôt un changement relatif à un état interne, et vous interprèterez sans doute ce fait curieux ainsi. Vous vous direz : « Cette sensation bizarre n'est pas provoquée par le mur, mais causée par ce neurochirurgien qui opère sur mon cerveau ». Mais qu'est-ce qu'une perception, sinon une sensation que l'on rapporte comme causée par un certain objet ou état de chose et qui est effectivement causée par cet objet ou cet état de chose ? Parfois, lorsqu'une sensation a une cause plausible dans l'environnement, on pense que c'est lui qui est responsable de cette sensation, mais cela pourrait être à tort. Un article[214] fait état d'une observation clinique curieuse. Alors qu'un neurochirurgien était en train de procéder à des stimulations de son cerveau en vue d'une opération pour cause d'épilepsie, une jeune fille de 16 ans fût soudain prise d'une crise de fou rire, lorsqu'il stimula une petite zone de 2 cm sur 2 cm. Intrigué, ce neurochirurgien procéda à divers tests et remarqua notamment qu'à chaque stimulation de la zone, la jeune fille, non seulement était prise de fou rire, mais

[214] Fried, I., Wilson, C. L., Macdonald, K. A. and Behnke, E. J. "Electric current stimulates laughter".

qu'en plus, elle assignait à un événement extérieur la cause de son rire. Autrement dit, son rire était accompagné d'une vive impression que sa cause réelle se situait dans l'environnement. Par exemple, lui présentant une image de cheval tout en stimulant la zone, elle rapporta, lorsqu'on lui demanda ce qui la faisait rire : « Le cheval est amusant ! ». Mais en réalité, la véritable cause de son rire était interne. Le fait que notre milieu interne reste assez constant par rapport à un environnement qui ne cesse de changer participe sans doute à l'illusion que nous sommes un observateur interne d'un spectacle externe, mais si c'était le milieu intérieur qui changeait le plus, et qu'il ne change pas de manière chaotique mais selon une certaine logique, n'aurions-nous pas la sensation que notre milieu interne est extérieur à nous ? Il semble bien que ce soit ce qui nous arrive chaque nuit. Lors du sommeil, les interactions avec l'extérieur sont très réduites car inhibées par le tronc cérébral, mais, du moins lors des phases de sommeil paradoxal, notre activité cérébrale est très riche, ce qui nous donne l'impression que nos rêves sont extérieurs à nous. Que l'on pense à présent au fait d'évoquer en nous un souvenir, une émotion, etc., ou même de boire de l'alcool ou du café. Dans chaque cas, on peut conceptualiser ce qui se passe comme une sorte d'action interne (dans les deux derniers cas, c'est une action interne indirecte puisqu'elle passe par des événements externes). Que l'on pense aussi au biofeedback. Cette technique permet, avec de l'entraînement, d'agir volontairement sur certains paramètres physiologiques habituellement hors de portée de notre volonté. Tout cela nous montre à quel point considérer une action comme étant forcément une action sur l'environnement est réducteur. Même les actions sur l'environnement ont toujours une finalité interne. N'est-ce pas la quête du bonheur qui, *in fine*, nous fait agir ? Or, le bonheur a une composante interne non négligeable : le sentiment de détente, de paix, de relaxation, que l'on éprouve après un effort méritant est bien évidemment corrélé à des états internes de l'organisme. Inversement, une action interne telle l'évocation d'un sentiment ou d'un

souvenir peut être un préparatif à une action plus efficace ou même constitutif de cette action. Par exemple, au théâtre, il est important de savoir évoquer en soi les émotions de son personnage, afin de bien le jouer. Un acteur entraîné passera par cette étape sans même avoir à y réfléchir, cela fera partie intégrante de son jeu de scène.

Nous avons présenté quelques intuitions internalistes qui conduisaient à penser que notre esprit est dans notre tête. Puis nous avons démonté ses intuitions en montrant que la vie mentale pouvait se concevoir entièrement comme rattachée à l'environnement. Pour être le plus juste possible, il convient sans doute d'adopter un troisième point de vue, que nous illustrerons par le schéma suivant :




Renverser les intuitions internalistes n'a pas pour but d'adopter une position située à l'extrême opposé, mais de mettre en évidence une certaine symétrie de la contribution de

l'environnement et de l'organisme à la perception et à l'action. Il faudrait donc plutôt voir chacun de ces aspects, extérieur et intérieur, non pas comme opposés dans une conceptualisation dichotomique et réductionniste du fonctionnement de l'esprit, mais au contraire comme complémentaires et profondément intégrés, comme constitutifs de la conscience ou de son émergence.

## 4. Cerveau planétaire et sociologie neuronale

Une autre objection pourrait prendre la forme suivante. Nous avons supposé, dans notre démonstration, que les interactions entre individus étaient au moins aussi riches et aussi constantes que celles ayant lieu entre les deux hémisphères d'un cerveau, mais peut-on légitimement affirmer une telle chose ? Peut-être pas, il est vrai, du moins pas dans des circonstances normales. Toutefois, même si la conscience interindividuelle qui devrait émerger des interactions entre individus n'était, le plus souvent, qu'embryonnaire, il semble exagéré, comme nous allons le voir, de la considérer comme totalement inexistante.

Cette seconde objection pose plus de problèmes, car il est vrai que, si nos hémisphères se désolidarisent rarement pour aller voir ailleurs ce qui se passe avec d'autres hémisphères, on ne peut pas en dire autant des relations interpersonnelles. Bien sûr, on peut avoir des bons amis, des gens que l'on trouve très sympathiques et avec lesquelles on communique très régulièrement. Mais il n'en demeure pas moins que ces rapports ne sont ni exclusifs ni constants. Cela a-t-il une importance ?

En fait, même si les hémisphères semblent relativement sages dans notre tête, il suffit de descendre d'un niveau pour se rendre compte de la jungle véritable que représentent les

interactions neuronales. Des connexions se créés et se défont régulièrement. C'est la plasticité cérébrale. Sans elle, point d'apprentissage. D'assez larges zones du cerveau peuvent ainsi se trouver réaffectées à une autre fonction, suite par exemple à une amputation ou un entraînement intensif. On sait même depuis quelques temps que les neurones se régénèrent. Toutes proportions gardées, il semble donc bien qu'à l'échelle des tissus cérébraux, les phénomènes de plasticité cérébrale ne sont pas un frein à l'émergence de la conscience, et sont même une condition nécessaire à sa propre plasticité. Il n'est donc pas clair que l'argument opposé aille dans le sens voulu. Il est même permis de supposer que la plasticité sociale pourrait jouer un rôle similaire, au niveau de la conscience interindividuelle, à la plasticité cérébrale, au niveau de la conscience individuelle.

En second lieu, on peut dire que certaines opérations neurochirurgicales, telles la commissurotomie, sont des traumatismes comparables à une rupture dans une relation sociale. Pourtant, il ne semble pas qu'une telle opération ait un effet dévastateur quant à la conscience de l'individu qui en est le siège. De même, il semble, intuitivement, que rétablir les connexions, ou ôter un des hémisphères et en mettre un autre à la place, ne détruirait pas la conscience. Bien sûr, cela la modifierait radicalement. Il n'est même pas dit que cela en préserverait l'identité numérique, mais on voit mal pourquoi l'individu résultant n'aurait plus du tout de conscience, s'il se comporte en tout point comme s'il était conscient. D'autres opérations neurochirurgicales, ainsi que des accidents cérébraux, peuvent altérer substantiellement la conscience, mais ne la supprime pas.

Enfin, que dire de la richesse des interactions sociales ? Est-elle comparable à ce qui se joue entre nos deux oreilles ? Cela n'a pas été le cas de tout temps. Cependant, il semble que cela soit de plus en plus vrai. L'Internet est une gigantesque toile où transitent des

milliards de bits d'information par seconde. Et nous n'en sommes encore qu'au début du développement de ce vaste réseau. L'analogie avec le cerveau est de plus en plus frappante, à tel point que l'on parle – de moins en moins au sens figuré – de mise en place d'un véritable cerveau planétaire. [215]

Dans un sens contraire, on peut considérer que la richesse de la communication inter-hémisphérique est sérieusement compromise par une commissurotomie, ce qui n'empêche pas, selon toute vraisemblance, aux patients split brain, après parfois un temps d'adaptation, d'être le siège d'une conscience unifiée dans la plupart des situations de la vie courante.

Concernant les questions de richesse et de constance des interactions, on peut donc envisager des cas, au niveau individuel, qui ressembleraient fortement à ce que l'on observe à un niveau social, et qui pourtant seraient, semble-t-il, compatibles avec la préservation d'une activité consciente. Ainsi, il n'est pas injustifié de maintenir notre analogie avec les groupes humains et la société et à affirmer prudemment l'existence, dans ces cas là aussi, d'une sorte de conscience, interindividuelle.

## 5. Conclusion

Nos conclusions, fondées sur l'hypothèse de l'esprit étendu, quant à l'existence possible d'une conscience interindividuelle, rejoignent, il nous faut le signaler, celles obtenues, d'une autre manière, par les sciences de la complexité. La systémique, la

---

[215] Cf. DE ROSNAY, Joël, *Le Macroscope*. Et aussi : HEYLIGHEN, Francis, JOSLYN, Cliff et TURCHIN, Valentin, *Principia Cybernetica Web* [en ligne] : < http://pespmc1.vub.ac.be/ >

cybernétique, l'étude des sociétés d'insectes, des phénomènes d'émergence, etc. conduisent au constat suivant : au sein des groupes d'individus émerge une conscience de groupe[216]. Celle-ci repose sur les interactions complexes qui sont entretenues au sein du groupe. Chaque individu est en quelque sorte un neurone d'un cerveau qui est le groupe lui-même. Et ce cerveau possède une conscience. Cette conscience, sans doute, est limitée, lorsque le groupe est simple et restreint. Mais, de plus en plus, notamment grâce aux possibilités offertes grâce à l'Internet, on assiste à l'émergence de ce que certains auteurs n'hésitent pas à appeler : le cerveau planétaire. Ce cerveau planétaire serait le soubassement matériel d'une véritable conscience planétaire, cette fois beaucoup plus complexe, et en constante évolution, grâce aux progrès constants de la technologie en matière de télécommunications. On pourrait peut-être même pousser cette considération un peu plus loin – quitte à nous livrer à quelque spéculation – jusqu'à l'Univers dans son entier car, malgré le vide intersidéral, des relations existent entre chaque partie de l'univers et chaque autre, entre chaque partie et l'univers entier, ne serait-ce que des relations spatiales de positionnement, de distance, etc. S'il existe une conscience ou du moins une sorte d'esprit de l'Univers (Dieu ?), il ou elle est sans doute, en l'état actuel des choses, rudimentaire, car on ne peut pas dire que les interactions interplanétaires – si elles existent ! – qui plus est avec d'hypothétiques êtres intelligents habitant quelque part ailleurs dans l'univers, soient foisonnantes ! Cependant, ce commencement d'esprit universel a peut-être une unité suffisante pour garantir l'identité de ses constituants, à la manière dont les *champs morphogénétiques* de l'embryon déterminent l'identité de ses cellules. Si l'univers est une sorte d'hypersphère finie, quoique illimitée, alors chaque élément qui le constitue peut être

---

216 Cf. DE ROSNAY, Joël, *Le Macroscope*. Et aussi : HEYLIGHEN, Francis, JOSLYN, Cliff et TURCHIN, Valentin, *Principia Cybernetica Web* [en ligne] : < http://pespmc1.vub.ac.be/ >

déterminé précisément par sa position en son sein, à l'aide, par exemple, de coordonnées polaires. Et ce, même si intrinsèquement, rien ne le différencie d'un autre élément.

Comme le souligne David Banach[217] :

> « Une part essentielle de votre identité est votre perspective, la somme totale des relations que vous entretenez avec le reste de l'univers, votre place dans le monde. La somme totale de ces relations a évidemment quelque effet sur les propriétés qui ont un lien avec vous, mais elle n'est pas réductible à ces propriétés. La perspective détermine l'ensemble de propriétés objectives qui comprend le contenu de cette perspective, mais elle n'est pas épuisée par ce contenu. La manière dont le contenu de chacune des relations individuelles qui constituent une perspective sera combiné pour faire un sujet unique n'est pas déterminée par leurs contenus individuels, mais par la manière dont elles sont combinées avec les autres relations qui constitue cette perspective. L'identité implique nécessairement, à la fois la perspective et un certain ensemble de propriétés objectives déterminées par cette perspective (appelons ces propriétés « essence »). Notons que la perspective implique nécessairement l'essence, mais l'essence en elle-même est universelle et ne peut pas déterminer un individu unique. Donc, un objet particulier requiert une perspective. Etre, c'est être une perspective (entre autres choses). »

Et même si cette considération est un peu trop audacieuse au regard de l'Univers dans son entier, elle est certainement justifiée, comme les sciences de la complexité et l'externalisme actif nous le suggère, au niveau planétaire. Ainsi, à la question que nous nous étions posée en introduction de cette partie : « Peut-on transposer ce modèle [les *champs morphogénétiques*] de la détermination de l'identité cellulaire lors du développement embryonnaire à la détermination de l'identité personnelle ? », nous pouvons répondre par

---

217 BANACH, David. "Who Do You Think You Are? Relations, Subjectivity, and the Identity of Persons".

l'affirmative. Il en découle que l'on peut comprendre l'identité des personnes au regard du rôle qu'elles occupent au sein de ce cerveau planétaire dont nous avons parlé.

David Banach poursuit plus loin :

> « L'identité à travers le temps implique **une continuité de perspective effectuée par l'action autonome**. Une chose est définie (partiellement) par une perspective qui détermine le contenu de cette chose et comment elle sera synthétisée dans un tout cohérent. A mesure que la chose change, c'est à la fois son contenu objectif et sa perspective qui change. La chose peut être dite rester la même chose si le déterminant de sa perspective dans la dernière phase est le contenu objectif synthétisé de la phase précédente, si sa perspective (la source de son identité présente) est causée par son essence passée prise comme un tout et non par des facteurs externes ou par quelque partie isolée de son essence passée. La continuité de notre identité est maintenue par l'action autonome, dans laquelle notre soi passé compris comme un tout, notre essence, détermine notre place dans le monde pour l'instant suivant, notre perspective (laquelle, en retour, déterminera ce que nous serons au moment suivant). Donc, l'essence passée détermine la perspective présente (au moins pour les choses qui ont une intégrité au cours du temps). Il semble que nous pouvons obtenir cela de deux manières : l'essence (passée) précède l'existence et l'existence précède l'essence (présente). En un sens différent, nous créons notre identité et en héritons à la fois. »

Ainsi, l'identité au cours du temps pourrait se comprendre de la façon suivante : à chaque instant, l'individu que nous sommes va causalement affecter la perspective que nous aurons à l'instant suivant – c'est-à-dire l'ensemble de relations que nous entretenons avec l'univers entier – et en retour, cette perspective va conditionner l'individu que nous serons à cet instant-là. Par exemple, l'individu Paul, qui entretient une certaine relation au reste de l'univers, et possède donc une certaine perspective, fait un pas en avant. Il s'ensuit qu'il modifie sa perspective (il se rapproche, mettons, d'un mètre d'une rivière, de deux mètres

d'une maison, etc.). Cette nouvelle perspective a une répercussion sur l'individu Paul (il voit à présent un clocher qui était caché derrière un immeuble, et apprend qu'il est, mettons, 18h32 ; cette expérience correspond à une certaine configuration neurobiologique différente de la précédente). L'individu Paul change donc qualitativement, entre ces deux instants. Mais dans la mesure où ce changement est lié à un changement de perspective, et dans la mesure ou ce changement de perspective est causé par l'individu Paul au premier instant, alors l'identité personnelle de Paul – qui est une continuation de perspective, et non d'individualité – est conservée entre ces deux instants.

Supposons que l'individu Paul, entre le premier et le deuxième instant, se soit dupliqué comme une amibe. Nous avons Paul1 qui effectue un pas en avant pendant que Paul2 fait un pas sur le côté. Néanmoins, Paul, avant cette duplication, avait une idée bien en tête : ou bien faire un pas en avant, ou bien faire un pas sur le côté. Il n'est pas possible qu'il ait pu fermement décider, à la fois de faire un pas en avant, et à la fois de faire un pas sur le côté. Dans ce cas-là, Paul s'identifie naturellement à la perspective de l'individu qui actualise son intention de départ.

Cela règle-t-il tous les problèmes liés à la duplication de personnes ? Peut-être pas. S'il est vrai que nos actions intentionnelles ne peuvent suivre, au même moment, deux chemins radicalement différents, et permettent donc de pointer valablement vers un et un seul individu, dans le cas du type de celui que nous avons envisagé, cela ne règle pas, semble-t-il, le problème de la duplication, lorsque aucune action intentionnelle n'est impliquée. Par exemple, Paul est dans son lit et dort. Pendant son sommeil, il subit une duplication en Paul1, qui est transporté jusqu'à la baignoire, et Paul 2, qui est transporté jusqu'au paillasson. L'intention première de Paul était de dormir dans son lit, ni dans la

baignoire, ni sur le paillasson. En ce cas, faut-il convenir que Paul est mort, parce que les nouvelles perspectives ne sont pas liées à la précédente, par l'intermédiaire d'une action autonome de l'individu Paul ? Et que penser du cas où Paul est un extraterrestre d'une espèce se reproduisant par division conservatrice de la structure microphysique de l'individu de départ et qui décide, à un moment donné, de se diviser en Paul1 et Paul2. ? Cette fois-ci, les deux nouvelles perspectives sont liées à l'ancienne par une action autonome de l'individu Paul, et l'on devrait donc dire, en toute logique, que Paul1 et Paul2 sont Paul. Mais comme, manifestement, Paul1 n'est pas Paul2 et que l'identité est une relation transitive, quelque chose ne colle pas.

La thèse de David Banach est intéressante en ce sens qu'elle suggère une piste plausible d'interprétation correcte des expériences de pensée de duplication. Néanmoins, elle ne semble fonctionner que dans un nombre restreint de cas. Les cas où une « action autonome » est impliquée mais n'est pas cause de la duplication.

### *D. Application de l'externalisme au problème de l'unité de la conscience*

La thèse de l'externalisme actif nous permet aussi de conceptualiser de manière satisfaisante un certain nombre de problèmes liés à l'unité de la conscience. Le problème de l'unité de la conscience est connexe à celui de l'identité personnelle. Dans les cas de fissions ou de fusions d'individus, l'unité de la conscience est mise en question, puisque l'on a, respectivement, une seule conscience qui en donne plusieurs et plusieurs qui en donnent une seule. Ce qui va nous intéresser plus particulièrement, ici, c'est le processus lui-même du passage de une conscience à plusieurs ou de plusieurs à une seule. Comment le

concevoir ? Est-ce d'ailleurs concevable ? Y a-t-il un moment précis où s'effectue le passage de une conscience à plusieurs, lors d'une fission qui suivrait un processus physique progressif ? Y a-t-il un moment précis ou s'effectue le passage de plusieurs conscience à une seule, lors d'une fusion qui suivrait un processus physique progressif ? Si oui, quels sont ces moments ? Lorsque la fission ou la fusion est encore incomplète, quelle est la nature de cette entité consciente dont les bases physiques sont partiellement mais pas entièrement dédoublées ? Comment s'y organisent les contenus expérientiels ? Les expériences sous-tendues par les états physiques localisés dans les parties non encore dédoublées ou déjà fusionnées, sont-elles tout simplement partagées par plusieurs centres de conscience, en dépit du fait que les contenus expérientiels sous-tendus par les états physiques localisés dans les parties déjà dédoublées ou pas encore fusionnées, ne sont pas partagés, réalisant de ce fait une sorte d'unité partielle de la conscience ? Les expériences se dédoublent-elles toutes, dès lors que la fission commence, et fusionnent-elles toutes, dès lors que la fusion se termine ? L'unité de la conscience réaliserait-t-elle la prouesse consistant à être indépendante de l'unité de son substrat ? Cela pourrait être en ce sens que, lors de la fusion ou de la fission de consciences, tous les contenus expérientiels impliqués seraient unifiés, dès lors qu'il existerait au moins un soubassement physique de la conscience en commun aux consciences fusionnant ou fissionnant. Cela pourrait être, au contraire, en ce sens que, lors de la fusion ou de la fission de consciences, les parties du soubassement physique de la conscience qui ne serait pas encore dédoublées ou qui seraient déjà fusionnées ne serait pas du tout le siège d'expériences conscientes, jusqu'à ce que le processus soit entièrement achevé.

Bien sûr, ces questions peuvent en partie concerner des cas d'individus partiellement dédoublés ou partiellement fusionnés qui ne sont pas engagés dans un processus de fission ou de fusion destiné à être complet.

Et concrètement, nous nous intéresserons au cas – actuel – des patients split brain, ces personnes dont on a sectionné le corps calleux afin de soulager des crises d'épilepsie particulièrement violentes, ainsi qu'à celui – actuel également – des jumeaux siamois *craniopagus*. Nous montrerons comment la thèse de l'externalisme actif nous aide à conceptualiser ces cas.

## 1. Les patients split brain

Le cas des patients callosotomisés (split brain) est d'un grand intérêt pour la métaphysique et l'ontologie de l'esprit. Premièrement, et avant toute chose, car ce cas est réel. Contrairement aux expériences de pensée, on ne peut donc pas reprocher aux philosophes qui se penchent sur ce cas de vouloir statuer sur le réel à partir de « fables ». Deuxièmement, car ce cas est des plus troublant pour le sens commun, défiant de nombreuses conceptions de l'identité personnelle ayant pour point commun de considérer qu'être une Personne, c'est être Une personne, et aussi une personne Une.

### a) Internalisme, substantialisme, et indétermination du nombre de centres conscients

Dans *Brain bisection and the unity of consciousness*, Thomas Nagel[218] s'attache à recenser les différentes interprétations classiquement données au curieux comportement des patients split-brain et à leur phénoménologie encore plus mystérieuse :

1. Les patients ont un seul esprit relativement normal associé avec l'hémisphère gauche, et les réponses émanant de l'hémisphère droit non verbal sont les réponses d'un automates, et ne sont pas produites par des processus mentaux conscients.

2. Les patients ont seulement un esprit, associé avec l'hémisphère gauche, mais il existe aussi (associé à l'hémisphère droit) des phénomènes mentaux conscients, non intégrés à un esprit, bien qu'ils puissent peut-être être attribués à l'organisme.

3. Les patients ont deux esprits, un qui peut parler et l'autre qui ne peut pas.

4. Ils ont un seul esprit, dont les contenus dérivent des deux hémisphères et sont plutôt bizarres et dissociés.

5. Ils ont un seul esprit normal la plupart du temps, tandis que les hémisphères fonctionnent en parallèle, mais deux esprits sont obtenus par les situations

[218] NAGEL, Thomas. "Brain Bisection and the Unity of Consciousness". *In:* PERRY, John. *Personal Identity*.

> expérimentales qui fournissent les résultats intéressants. (Peut-être que l'unique esprit se divise en deux et se reforme après la fin de l'expérience.)

Il montre qu'aucune interprétation n'est réellement satisfaisante :

Tout d'abord, les options 1 et 2 ne rendent pas compte du caractère très complexe, marqué d'intentionnalité, d'intelligence et de rationalité du comportement qui prend naissance dans l'hémisphère droit. Même s'il lui manque le langage, il assez problématique de lui dénier toute conscience. Ensuite, les actions initiées par l'hémisphère droit jouissent d'une bonne intégration et d'une bonne exécution, elles ne sont ni vagues ni hésitantes. Par conséquent, on ne peut pas parler non plus de phénomènes mentaux conscients mais qui seraient isolés, non intégrés à un esprit.

L'argument généralement évoqué à l'appui du caractère inconscient de ce qui se passe dans l'hémisphère droit repose sur le témoignage des patients. Mais sachant que les centres du langage sont généralement présents dans l'hémisphère gauche, et que ce dernier, chez les patients split brain, ne peut évidemment pas savoir ce qui se passe dans l'hémisphère droit, l'argument n'est pas recevable !

Qui plus est (et malheureusement, Thomas Nagel n'en parle pas dans son article), signalons qu'il existe des cas de non-latéralisation du langage, c'est-à-dire que chez certains individus, les centres du langages se trouvent présent un peu dans chaque hémisphère. Plus rarement encore, il est arrivé que de tels individus se soient trouvés à subir une commissurotomie. Dans ce cas, il est possible de recueillir un témoignage verbal de la part **des deux** hémisphères. Ce témoignage, pouvant présenter d'importantes discordances

selon que l'on interroge l'un ou l'autre hémisphère, démontre, non seulement la possibilité de l'existence d'un esprit associé à l'hémisphère droit, mais suggère également que cet esprit serait distinct de celui associé à l'hémisphère gauche. On aurait donc deux esprits chez ces individus (option 3).

Si cette option est plausible relativement aux expériences menées en laboratoire sur ces patients, elle échoue à rendre compte du caractère pourtant hautement intégré de leur comportement dans les situations de la vie ordinaire, où les stimuli de l'environnement sont quasiment les mêmes pour l'un ou l'autre hémisphère, ce qui permet un traitement en parallèle assez concordant. Bien sûr, l'absence d'un corps calleux intact limite drastiquement la possibilité d'une intégration s'effectuant sur un plan cérébral. Les expériences de ces individus sont, si on peut dire, intégrées dans l'environnement lui-même, ou par le biais de stimulations ipsilatérales. Mais ce mode original d'intégration doit-il nous faire douter de la présence d'un seul esprit chez ces gens, au moins dans les situations de la vie quotidienne ? Pour Thomas Nagel, il est étrange de suggérer que nous ne pouvons pas attribuer toutes ces expériences à la même personne, juste à cause de quelques étrangetés concernant la manière dont l'intégration est effectuée.

D'où l'option 4, qui peut sembler plus adéquate dans la majorité des cas de la vie quotidienne. Néanmoins, si nous attribuons l'intégration à un seul esprit, nous devons aussi lui attribuer la dissociation expérimentalement évoquée, et cela n'est pas simple ! En effet, cette dissociation, lorsqu'elle se produit, est véritablement très profonde, si bien que l'on peut se demander quel socle commun pourrait répondre à une définition minimale du mot « esprit ». Et si un tel socle n'existe pas, c'est donc que l'on ne peut pas parler d'un seul esprit, mais de deux, ceux-ci se trouvant, en effet, en « aval » de la dissociation.

Comme le note Thomas Nagel, l'option 4 :

> « rend difficile de concevoir ce que cela fait d'être une de ces personnes [callotomisées]. L'absence d'interaction au niveau d'un système de contrôle préconscient serait compréhensible. Mais l'absence d'interaction dans le domaine de l'expérience visuelle et de l'intention consciente malmène nos présupposés concernant l'unité de la conscience, lesquels sont fondamentaux pour notre compréhension d'un autre individu comme personne. Ces présupposés sont associés à notre conception de nous-mêmes, laquelle, jusqu'à un point considérable, conditionne notre compréhension des autres. »

Un peu plus loin, Thomas Nagel explicite ces préconceptions :

> « Grossièrement, nous supposons qu'un esprit singulier a suffisamment d'accès immédiat à ses propres états conscients pour que, pour les éléments de l'expérience ou les autres événements mentaux qui se produisent simultanément ou dans une proximité temporelle rapprochée, l'esprit qui est leur sujet peut aussi expérimenter les *relations* plus simples entre eux, s'il fait un effort en ce sens. Donc, nous supposons que lorsqu'une personne singulière a deux impressions visuelles, il peut généralement aussi expérimenter la ressemblance ou la différence de leur coloration, forme, taille, la relation de leur position et mouvement à travers son champ visuel, etc. La même chose peut être dite des connexions entre modalités sensorielles. Les expériences d'une personne singulière sont supposées prendre place dans un domaine expérimentalement connecté, de manière à ce que les relations entre les expériences puissent être substantiellement capturées en expériences de ces relations. »

En note, il indique :

> « Un type de relation peut être maintenu entre des éléments dans l'expérience d'une personne singulière qui ne peut pas être maintenu entre les éléments de l'expérience de personnes distinctes : sembler similaire en couleur, par exemple. Dans la mesure où notre concept de

> similarité de l'expérience dans le cas d'une personne singulière est dépendant de son expérience de la similarité, le concept n'est pas applicable entre personnes. »

On pourrait vouloir remettre en cause notre concept traditionnel de personne, mais comment, dès lors, faire la distinction entre un individu dont l'esprit serait complètement dissocié et un individu qui possèderait deux esprits ? Mieux vaut s'en tenir à notre concept habituel de personne, qui a nécessairement un sens réel, du moment qu'on le prend pour synonyme de domaine conscient unifié incarné. Aussi, selon cette définition, il y aura autant de personnes qu'il y aura de tels domaines. Nous revoilà avec l'option 3. Mais les critiques qui lui ont été faites restent valables. Si le mot « personne » est synonyme de « domaine conscient unifié incarné », alors il y a nécessairement une seule personne chez les patients split brain lorsque l'on n'observe pas de dissociation, c'est-à-dire dans la majorité des situations de la vie courante.

L'option 5 peut sembler être un bon moyen de régler le dilemme : il y a un seul esprit, chez les patients split brain, dans les situations ordinaires, et cet esprit se dédouble lorsque ces patients sont placés dans certaines conditions expérimentales.

Toutefois, Thomas Nagel estime que cette option est également intenable. Ses raisons sont les suivantes :

- L'hypothèse serait entièrement **ad hoc** : il proposerait d'expliquer un changement dans les termes d'un second changement, sans suggérer aucune explication de ce second changement

- Il n'y aurait rien, concernant la situation expérimentale, dont on pourrait espérer qu'il produise un changement **interne** fondamental chez le patient.

- En fait, aucun changement **anatomique** n'est produit

- L'idée d'un esprit qui surgit dans l'existence et s'évanouit dans le néant subitement est assez étrange et mériterait une explication qui aille plus loin que le constat de son adéquation à expliquer le comportement des patients callosotomisés.

- Mais en fait, les données comportementales ne seraient même pas expliquées par cette hypothèse, simplement parce les réponses intégrées des patients et leurs réponses dissociées ne sont pas clairement séparées dans le temps. Durant le temps des expériences, le patient fonctionne largement comme s'il était un individu singulier : dans sa posture, dans le suivi des instructions sur où fixer son regard, dans tout l'éventail de contrôle comportemental trivial impliqué dans le fait de se situer lui-même en relation à l'expérimentateur et à l'appareillage expérimental. Les deux moitiés de son cerveau coopèrent complètement exceptées au regard de ces « inputs » très spéciaux qui les atteignent séparément et différemment.

De l'échec de ces cinq interprétations à rendre compte de la condition des patients callosotomisés, Thomas Nagel en conclut prudemment à l'inadéquation pure et simple du concept de personne ou d'esprit – du moins tel qu'on le définit traditionnellement – à

rendre compte de la réalité de la condition des patients split brain, car il n'y a pas un **nombre entier** d'esprits individuels que ces patients peuvent être dits posséder.

*« Cela va à contre courant,* nous dit Thomas Nagel, *de notre besoin de construire les états mentaux que nous attribuons à autrui sur le modèle des nôtres. Quelque chose, dans la conception ordinaire de la personne, ou dans la conception ordinaire de l'expérience, nous conduit à rechercher un compte-rendu de ces cas que cette même conception nous rend impossible à obtenir. »* Selon une autre perspective, on peut cependant comprendre que ce qui diffère fondamentalement de l'ordinaire ne puisse pas spécialement être décrit en des termes ordinaires… Thomas Nagel nous invite même, sur la base de ces cas extraordinaires que sont les patients split brain, à une remise en question plus générale du concept de sujet singulier de la conscience lorsqu'il s'applique à nous-même.

En effet, même les individus ordinaires n'ont pas une conscience toujours parfaitement unifiée, loin s'en faut. Il y a illusion d'unité, et cette illusion serait de nature linguistique. Ce serait l'usage de la première personne du singulier qui nous conduirait à nous considérer comme des êtres uniques et unifiés :

> « L'illusion consiste dans la projection vers l'intérieur, au centre de l'esprit, du sujet même dont nous essayons d'expliquer l'unité : la personne individuelle avec toutes ses complexités. Le compte rendu ultime de ce que nous appelons un esprit singulier consiste en une énumération des types d'intégration fonctionnelle qui le caractérise. Nous savons qu'elles peuvent être érodées de différentes manières, et à différents degrés. La croyance selon laquelle même dans leur version complète elles peuvent être expliquées par la présence d'un sujet numériquement singulier est une illusion. Ou bien ce sujet contient la vie mentale, dans ce cas il est complexe et son unité doit être expliquée par

> l'opération unifiée de ses composants et fonctions, ou bien c'est un point sans extension, dans quel cas il n'explique rien du tout. »

Nous nous permettrons toutefois de formuler quelques remarques critiques à l'encontre de certains points de la démonstration de Thomas Nagel. Nous n'avons pas grand'chose à objecter à sa critique des options 1 à 4. Cependant, nous estimons que sa critique de l'option 5 et ses conclusions souffrent de certaines faiblesses. Nous pensons aussi que l'option 5 peut très bien se concilier avec nos intuitions sur l'identité personnelle.

Lorsque Thomas Nagel remet en question notre conception traditionnelle de la personne comme unifiée sous prétexte que la réalité nous montrerait une personne plus ou moins fractionnée, au lieu de nous laisser troubler par le paradoxe apparent, nous pouvons aussi réagir de la façon suivante : il n'est possible d'admettre un fragmentation et une défragmentation de la personne (ou de l'esprit) que dans la mesure ou l'unité de la conscience n'est pas une caractéristique essentielle de ce concept. Or, il paraît judicieux d'affirmer au contraire qu'elle l'est. Sinon, comment différencierions-nous des états mentaux conscients dissociés d'une même personne des états mentaux conscients de deux personnes distinctes ? Par conséquent, selon cette optique, la fragmentation d'une personne ne résulte pas en une personne fragmentée, pas plus qu'en une non-personne, mais bien en autant de personnes qu'il y a de fragments. Et la réunification de ces fragments, n'implique pas la défragmentation d'une personne ou l'apparition d'une personne à partir d'une non-personne, mais la fusion de plusieurs personnes (autant que de fragments) en une seule.

Il est bien sûr difficile de rendre intelligible cela si on en reste à un certain nombre de présupposés ontologiques, ceux-là mêmes que l'on voit poindre en arrière-plan des objections de Thomas Nagel à l'option 5. Ces présupposés ontologiques sont les suivants :

- L'internalisme

- Le substantialisme

L'internalisme : suivant cette conception, l'esprit est entièrement contenu dans le cerveau, ce qui rend effectivement assez inenvisageable l'idée que l'on puisse passer d'un esprit à deux puis à nouveau à un, par le seul jeu des conditions environnementales. Pour la conception externaliste, en revanche, il n'y a pas un tel problème : les conditions environnementales peuvent dupliquer ou fusionner l'esprit car elles font partie intégrante de l'esprit.

Le substantialisme : lorsqu'un prestidigitateur cache un lapin avec un voile, puis qu'il le retire et qu'il y a deux lapins identiques au premier lapin, on peut légitimement se demander : d'où vient le deuxième lapin ? Il n'est pas concevable que le lapin de départ se soit divisé pour en donner deux, et ce, pas seulement pour des raisons biologiques, mais aussi pour des raisons plus profondes. Il n'est pas concevable non plus qu'un des lapins soit apparu *ex nihilo*, et ce pour des raisons physiques certes, mais aussi pour des raisons logiques et métaphysiques. « Rien ne se perd, rien ne se créé, tout se transforme », affirmait Lavoisier. Cette loi, qui est (de moins en moins) valable en physique, a son équivalent en logique et en métaphysique. Elle postule que la quantité de substance, dans l'univers, est nécessairement constante au cours du temps. Si un seul lapin pouvait en donner deux par division, sans apport extérieur de substance, cela signifierait que de temps en temps au

moins, il serait vrai que 1=2 ![219] De plus, nous n'avons aucune idée du processus effectif par lequel une telle division multiplicative pourrait avoir lieu. Cet apport supplémentaire de substance ne peut pas davantage venir d'une création *ex nihilo*, car cela signifierait que de temps en temps au moins, il serait vrai que 0=1 ! « L'Être est, le Non-Être n'est pas », disait Parménide. Il nous est particulièrement difficile d'imaginer une création *ex nihilo*, car on n'a aucune idée du processus effectif par lequel une telle création pourrait se réaliser[220]. La seule manière dont nous créons et pouvons créer est par un déplacement d'éléments préexistants. Si donc on voit d'abord un lapin, puis deux lapins identiques au premier, on en conclura nécessairement qu'au moins un des deux lapins vient d'ailleurs (mais d'où ?), qu'il n'est pas le même que le lapin de départ. **Il a fallu un apport de substance**. Si le prestidigitateur recouvre ces deux lapins d'un voile, puis qu'il n'y a plus qu'un seul lapin, on peut se demander : où est passé un des lapins ? Il n'est pas concevable non plus que les deux lapins aient fusionné pour n'en donner plus qu'un ou qu'un des lapins ait tout

[219] Une autre solution serait de dire qu'il y a déjà deux lapins au départ, mais qu'ils sont indiscernables et localisés dans le même espace-temps. Dans une ontologie substantialiste, on voit mal comment cela serait possible. Là où il y a de l'Être, il n'y a aucune place pour du Non-Être que l'on pourrait combler. De plus, comment distinguer métaphysiquement deux Êtres qui seraient physiquement indiscernables et qui occuperaient le même espace en même temps ?

[220] D'un point de vue métaphysique, on peut énoncer le principe général suivant : **tout effet est contenu dans sa cause ou égale à celle-ci**. En effet, si un être quelconque existe, alors de deux choses l'une : ou bien cet être a commencé à exister, ou bien il a toujours existé. Si un être existe, c'est que c'est être est possible. S'il a commencé à exister, alors de deux choses l'unes : ou bien il a toujours été possible, ou bien sa possibilité a commencé à un certain moment. La deuxième option est paradoxale : si un être est possible, c'est qu'il est possible qu'il soit possible, ou bien qu'il est nécessaire qu'il soit possible. La première option nous renvoie à une régression à l'infini, la seconde s'analyse simplement en : cet être a toujours été possible. Mais comment un être peut-il avoir toujours été possible ? Il y a trois possibilités : ou bien cet être est en fait contenu dans sa cause, ou bien il dépasse celle-ci, ou bien il est sa propre cause (autrement dit, il n'a pas de cause), ce qui revient à affirmer qu'il est égal à sa cause. Enfin, si l'effet dépasse sa cause, ce qui dépasse la cause, dans l'effet, peut être : contenu dans une autre cause, dépasser cette deuxième cause, être sa propre cause. Le second cas nous conduit à une régression à l'infini. Nous retiendrons donc les deux autres.

simplement disparu dans le néant, et pas simplement pour des raisons biologiques ou physiques, mais pour les mêmes raisons logiques et métaphysiques. Il nous est particulièrement difficile d'imaginer une annihilation, car on n'a aucune idée du processus effectif par lequel une telle annihilation pourrait se réaliser. Il nous est particulièrement difficile d'imaginer une fusion qui n'augmente pas la quantité de substance, car nous n'avons aucune idée du processus effectif par lequel cela serait possible. La seule manière dont nous détruisons et pouvons détruire est par un déplacement d'éléments préexistants (nous détruisons en fait par un processus similaire à la manière dont nous créons). Si donc on voit d'abord deux lapins identiques, puis deux lapins identiques au premier, on en conclura nécessairement qu'au moins un des deux lapins est parti ailleurs (mais où ?), qu'il ne s'est pas fondu en l'autre, et qu'il n'a pas non plus été annihilé. **Il a fallu un retranchement de substance**. Ce raisonnement, somme toute valable pour les lapins en l'état actuel de nos connaissances, peut-il s'appliquer à la conscience ? Si on dit que oui, alors la présence de deux consciences chez les patients split brain lors de certaines situations expérimentales est incompréhensible. Bien sûr, sur le plan de la matière, on comprend bien comment un cerveau peut se transformer en deux demi cerveaux. Mais sur le plan de la conscience, on ne comprend pas comment une conscience peut se transformer en, non pas deux demi-consciences (le terme n'a aucun sens), mais en deux consciences. D'où vient au moins une de ces deux consciences ? Bien que les progrès des sciences cognitives aient relégué Descartes dans le « Musée des Philosophies Périmées », son substantialisme de la conscience est encore très vivace. Bien sûr, on ne conçoit plus la *res cogitans* comme séparée de la *res extensa* (la première est désormais une certaine fraction de la seconde), mais on n'a pas abandonné la *res cogitans* pour autant. Aujourd'hui, la *res cogitans*, c'est tout simplement l'unité **anatomique, structurelle,** de la substance qui constitue le cerveau. Mais si on dit que ce raisonnement n'est pas valable, alors il y a peut-être un espoir

d'intelligibilité. Il nous faut troquer notre ontologie substantialiste pour une ontologie dynamique du process, de l'événement. Être une personne, c'est être, non pas un morceau de cerveau, mais être une unité de **fonctionnement dynamique**. Un processus peut-il donner deux processus identiques sans apport de substance ? Deux processus peuvent-ils donner un seul processus, sans retranchement de substance ? Deux processus numériquement distincts mais qualitativement identiques peuvent-ils coexister au même endroit en même temps ? Mais d'abord, qu'est-ce qu'un processus ? D'après la *Standford Encyclopedia of Philosophy*[221] :

> « La philosophie du processus place les processus au premier rang des préoccupations philosophiques et spécifiquement ontologiques. Le processus doit être ici analysé de façon sensiblement usuelle – comme *une séquence structurée séquentiellement de phases ou de stages successifs*. Trois facteurs viennent en conséquence au premier plan :
>
> 1. Qu'un processus est un complexe – une unité de phases ou de stages distincts. Un processus est toujours une question de maintenant ceci, maintenant cela.
> 2. Que ce complexe a une certaine cohérence temporelle et unité, et que les processus en conséquence ont une dimension temporelle inéliminable.
> 3. Qu'un processus a une structure, un format générique formel en vertu duquel chaque processus concret est muni d'une forme ou d'un format. »

Bref, une ontologie du processus, par rapport à une ontologie substantialiste, met en exergue le changement, par rapport à l'immuabilité. Il nous faut ajouter également que la notion de processus, au contraire de celle de procédure, présuppose une certaine liberté. Alors qu'une procédure est rigide, un processus peut faire une large place à l'improvisation.

---

[221] RESCHER, Nicholas. "Process Philosophy". *In: Stanford Encyclopedia of Philosophy*.

Nous pouvons, à présent, répondre aux autres questions :

- Un processus peut-il donner deux processus identiques sans apport de substance ? Oui, et pour cela, il suffit qu'après une certaine étape du processus de départ, nous ayons deux étapes en parallèle. Nous aurons alors deux processus dont on pourra dire que les premières étapes sont communes jusqu'à un certain niveau. Il est important de bien comprendre que, d'un point de vue sémantique, ces deux processus ne commencent pas au moment où ils divergent, mais lorsque ce que nous avons appelé le « processus de départ » commence. En effet, la divergence redéfinit rétrospectivement ce que nous croyions être un seul processus (c'est-à-dire les premières étapes) comme ayant été, en fait, deux processus coexistants. Le point 2 de la définition de la *Standford Encyclopedia of Philosophy* est très clair à ce sujet : un processus a « une certaine cohérence temporelle et unité ». Il y a ce que nous pouvons considérer être un *externalisme temporel*[222] : le nombre de processus en jeu à un temps t, dépend du nombre de processus discernables générés par « branchement » *après* ce temps t. Le futur a une action causale sur le passé. Ce n'est évidemment pas une causalité efficiente, c'est une causalité téléologique.

- Deux processus peuvent-ils donner un seul processus, sans retranchement de substance ? Oui. Inversement, si des processus distincts présentent, à un moment, une étape commune, alors, d'une certaine manière, ils ne formeront plus qu'un seul processus. D'une certaine manière seulement :

222 Cf. STONEHAM, Tom. "Temporal Externalism"

pour un observateur qui n'a pas vu ce qui s'est passé avant, il n'y aura qu'un seul processus. Mais d'un point de vue sémantique, nous devons admettre qu'il y a deux processus en jeu, dont les étapes se confondent, à partir d'un certain point. Nous avons aussi, ici, un externalisme temporel, curieusement beaucoup plus facile à admettre, car il va dans le sens « normal » : du passé vers le futur.

- Deux processus numériquement distincts mais qualitativement identiques peuvent-ils coexister au même endroit en même temps ? La réponse aux deux précédents points nous permet de répondre par l'affirmative à ce dernier point.

Quelques difficultés subsistent cependant : on peut admettre, à la limite, une rétrocausalité dans le cas des procédures, où le déterminisme strict du système permet de savoir, déjà, à l'avance, le nombre de procédures existantes. Dans le cas des processus non strictement déterminés, un tel savoir est impossible et cette impossibilité est souvent considérée comme une condition nécessaire pour que l'on puisse parler d'indétermination, de liberté, d'improvisation. C'est une ignorance essentielle. Pourtant, si l'on doit admettre cette hypothèse de rétrocausalité, alors on doit admettre que d'une certaine manière, l'Univers « sait » déjà ce qui va se passer, ce qui paraît incompatible avec un indéterminisme.

Dans notre partie sur les critères biologiques et psychologiques de l'identité personnelle, nous avons proposé quelques pistes de réflexion, plus ou moins adéquates. Grosso modo, il s'agit de reconnaître une sorte de dépendance sortale du processus

« conscience », « esprit » ou « personne » par rapport au processus « organisme » ou « individu ». Une même conscience (donc un même esprit, une même personne) ne peut exister que dans un même organisme ou individu. Mais l'inverse n'est pas vrai : un même organisme ou individu peut instancier (et même : instancit) plusieurs consciences (ou esprits ou personnes) numériquement distinctes. Cela ne répond pas, toutefois, à toutes les objections.

C'est peut-être à ce moment là que l'on peut ressortir la thèse de David Banach, dont nous avons montré qu'elle décrivait, de manière satisfaisante, les cas de duplication où une « action autonome » est impliquée. Or l'usage de notre libre-arbitre (s'il existe) fait sans conteste partie de ces « actions autonomes ». Cependant, encore une fois, cela ne règle pas tous les problèmes, car l'indéterminisme partiel qui caractérise supposément l'avenir d'une personne ne se résume pas à l'effet de son supposé libre-arbitre.

### b) Externalisme, ontologie du processus et conscience partiellement unifiée

Admettons cependant ce nouveau paradigme, dont les piliers sont :

- L'externalisme

- L'ontologie du processus

Nous pouvons appliquer ce paradigme pour réinterpréter à sa lumière le cas des patients split brain. C'est ce que fait Susan Hurley dans son article : "Action, the Unity of Consciousness and Vehicle Externalism"[223].

Ce qui garantit l'intelligibilité de ces cas réside dans un nouveau concept : l'unité partielle (de la conscience). Mais ce concept, lui-même, n'est intelligible que si on a accepté de renoncer à une conception internaliste et substantialiste de la conscience.

Qu'est-ce que cette unité partielle de la conscience ? Ecoutons ce qu'en dit Susan Hurley :

> « L'unité totale [de la conscience] implique que si deux états conscients sont co-conscients à un moment, alors chacun est aussi co-conscient avec tous les états avec lequel l'autre est co-conscient à ce moment. C'est-à-dire, une pleine unité implique que …*est co-conscient et simultané avec*… est une relation transitive entre états conscients. Dans une conscience partiellement unifiée, deux états qui ne sont pas co-conscients l'un avec l'autre peuvent néanmoins être tout deux co-conscients avec le même troisième état. »[224]

Un problème se pose : qu'est-ce que cela veut dire, subjectivement, d'avoir une conscience unifiée ? Quel effet cela fait ? Nous avons déjà donné un élément de réponse, lors de notre analyse du traitement, par Thomas Nagel, des options 4 et 5. Comme on peut s'y attendre, cela ne fait rien de spécial. « Pour imaginer une conscience partiellement unifiée, nous devons imaginer quelque chose non seulement au sujet des types et contenus

---

[223] HURLEY, Susan. "Action, the unity of consciousness, and vehicle externalism". *In:* CLEEREMANS, Axel. *The Unity of Consciousness: Binding, Integration, Dissociation.*

[224] HURLEY, Susan. "Action, the unity of consciousness, and vehicle externalism", p. 1

des expériences, mais aussi quelque chose au sujet de leur identité », nous dit Susan Hurley. « C'est certainement difficile d'imaginer à quoi cela peut ressembler. Mais c'est parce qu'il n'y a pas d'effet particulier que cela fait, pour des expériences avec le même contenu, d'être identiques ou différentes. ».[225]

L'approche en « what it is like » chère à Thomas Nagel doit donc être rejetée ici. Elle n'est pas en mesure de nous aider à comprendre l'unité partielle.

Alors comment approcher concrètement cette unité partielle ? Avec nos fameux patients split brain, justement.

Et nous revoici avec l'option 5. Rappelons-là pour mémoire :

Selon l'option 5 envisagée par Thomas Nagel, les patients split brain ont un seul esprit normal la plupart du temps, tandis que les hémisphères fonctionnent en parallèle, mais deux esprits sont obtenus, dès lors qu'une situation expérimentale est mise en place, dans laquelle les stimuli arrivant à l'hémisphère gauche et ceux arrivant à l'hémisphère droit sont différents.

Susan Hurley prend acte du fait que l'on ne peut établir un isomorphisme entre la structure de la conscience et la structure neuronale. Pour elle, c'est de facteurs moteurs que dépend l'unité de la conscience plus ou moins complète des patients split brain (et même des sujets normaux). Autrement dit, ce qui va déclencher les divisions/réunifications plus

[225] Ici, Susan Hurley, bien qu'en procédant différemment, rejoint en substance Wittgenstein, sur l'impossibilité d'identifier des états internes privés.

ou moins prononcées de la conscience, cela va être des actions motrices, et même plus précisément des intentions motrices.

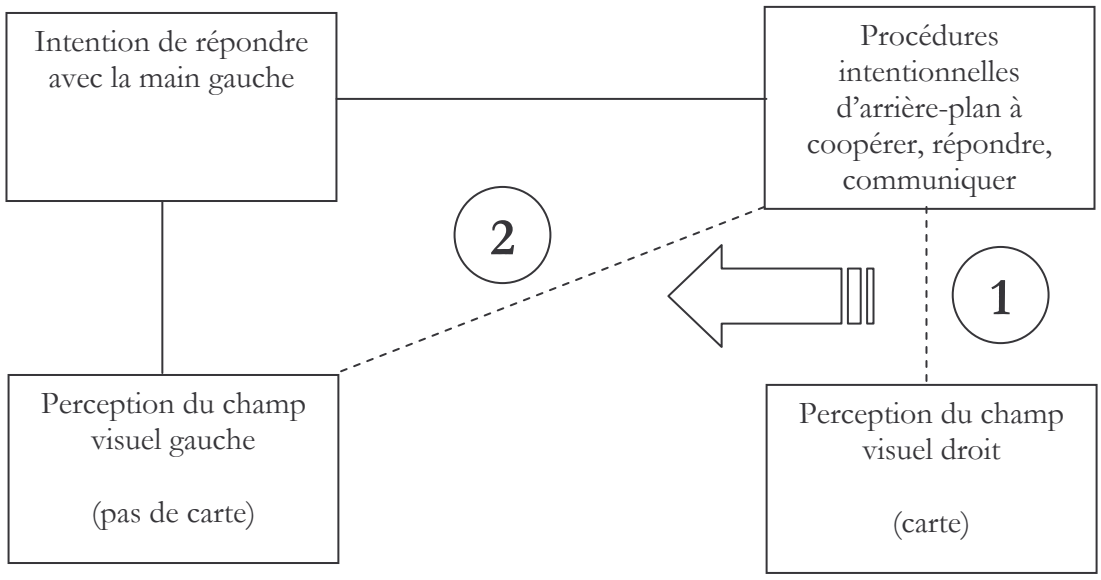
Avant de développer cette thèse, Susan Hurley s'attache tout d'abord à démolir la présupposition d'isomorphisme, par l'analyse du cas d'un patient split brain étudié par Trevarthen[226].

> « Le patient de Trevarthen essayait de réaliser une tâche de la main gauche relative à un objet [une carte] situé dans le champ visuel droit pendant qu'il fixait un point central. La fixation est supposée garantir que l'information de chaque moitié du champ visuel atteint seulement l'hémisphère contralatéral. Trevarthen rapporte que le patient dans ce cas
>
> > "… ne pouvait répondre et décrivait la forme comme s'évanouissant immédiatement comme le mouvement commençait… Ici le témoignage du patient était que l'image de l'objet, initialement vue, sortait de la conscience de l'hémisphère gauche à l'instant où un mouvement, initié par l'hémisphère droit, commençait… Tandis que le sujet se préparait à répondre, fixant attentivement… [l'objet] était invisible, même quand il était déplacé, jusqu'à l'instant où il croisait le méridien vertical. L'apparence du coin de la carte passant la ligne médiane dans le champ visuel gauche déclenchait une réponse rapide et puissante avec la main gauche en attente" (Trevarthen 1984, p. 333) »

Que se passe-t-il ? Nous pouvons mieux le comprendre avec le schéma suivant :

---

[226] TREVARTHEN, Colwyn, (1984) "Biodynamic Structures, Cognitive Correlates of Motive Sets and the Development of Motives in Infants", p. 333. *In* : Prinz, Wolfgang and SAUNDERS, A.F., *Cognition and Motor Processes*. Voir aussi : TREVARTHEN, Colwyn and SPERRY, Roger Wolcott. "Perceptual Unity of Ambient Visual Field in Human Commissurotomy Patients". Ainsi que : TREVARTHEN, Colwyn. "Analysis of Cerebral Activities that Generate and Regulated Consciousness in Commissurotomy Patients". *In:* DIAMOND, Stuart J. and BEAUMONT, J. Graham. *Hemisphere Function in the Human Brain*. Et : TREVARTHEN, Colwyn. "Modes of Perceiving and Modes of Acting". *In:* PICK, Herbert L., and SALTZMAN, Elliot (Eds.). *Modes of Perceiving and Processing Information*. Toutes ces références étant citées par Hurley dans son article.




Légende :

——— Co-conscience (permanente)

- - - - - - Co-conscience (temporaire)

① Avant l'initiation du mouvement

② Après l'initiation du mouvement

Voyons à présent pourquoi l'on ne peut retenir la thèse d'un isomorphisme pur et simple entre la structure de la conscience et les structures neuronales.

Tout d'abord, les mécanismes sous-corticaux ne supportent normalement pas la conscience (bien qu'ils traitent l'information). Or, la commissurotomie crée une dissociation à un niveau cortical, laissant l'unité sous-corticale intacte. Elle devrait donc créer, non pas une conscience partiellement unifiée, mais une conscience complètement dissociée. Ce n'est pas ce que l'on observe.

Une objection, plus philosophique, est la suivante : Dennett[227] et Millikan[228] (1993) nous font prendre conscience qu'il faut distinguer entre :

- Les propriétés *représentées* dans le contenu de la conscience (par exemple, le vert de l'herbe)

- Les propriétés des *véhicules* (biologiques) du contenu de la conscience (par exemple, le vert de l'herbe n'est pas représenté par des états cérébraux eux-mêmes verts…)

On peut étendre ces considérations au sujet du contenu de la conscience à sa structure. Ainsi, une conscience unifiée, partiellement unifiée, et dissociée, n'a pas nécessairement besoin d'être supportée par une structure cérébrale respectivement unifiée, partiellement unifié, et dissociée. Il n'y a pas un isomorphisme obligatoire entre la structure de la conscience et l'anatomie du cerveau.

Des observations cliniques sont à l'appui de cette considération. Ainsi, il arrive que des personnes naissent sans corps calleux – cette anomalie est qualifiée de *callosal agenesis* et on parle alors de sujets acallosaux. Il est arrivé aussi que des enfants victimes précocement de violentes crises d'épilepsie aient à subir une commissurotomie. Dans les deux cas, on remarque que, en l'absence de corps calleux, ils présentent néanmoins une remarquable unité de la conscience dans les situations ordinaires, et même, en général, expérimentales,

---

227 DENNETT, Daniel Clement, *Consciousness Explained*.

228 MILLIKAN, Ruth Garrett. "Content and Vehicle". *In:* EILAN, Naomi., MCCARTHY, Rosaleen and BREWER, Bill (Eds.). Spatial Representation

comparable à celle de sujets « normaux ». Inversement, des gens ayant un corps calleux intact peuvent présenter une conscience dissociée. Soit par alternance, comme chez les victimes du syndrome de personnalité multiple, soit en simultané, comme chez les schizophrènes (unité partielle avec « émiettement » des parties dissociées de la conscience[229]).

Nous pouvons récapituler ces observations dans le tableau suivant :

| | Corps calleux intact | Corps calleux absent ou lésé |
|---|---|---|
| Conscience unifiée | Sujets normaux | Sujets acallosaux / Sujets ayant subis précocement une commissurotomie |
| Conscience au moins partiellement dissociée | Sujets victime d'un syndrome de personnalité multiple / Schizophrènes | Sujets ayant subis une commissurotomie à l'âge adulte |

Chez les sujets acallosaux ou les patients split brain ayant subi une commissurotomie durant l'enfance, il se met en place une compensation. Celle-ci peut prendre deux formes :

[229] Par cet « émiettement » nous voulons signifier qu'il n'y a pas, chez les schizophrènes, une individualisation franche et nette de deux consciences partiellement dissociées et partiellement unifiées. Il y a plutôt un noyau conscient central, co-conscient avec une pléthore d'états conscients isolés, non co-conscients entre eux. Ainsi, les schizophrènes pourront donc se trouver à aimer et détester à la fois la même personne, ou à ressentir la tristesse et exprimer la joie, etc.

- Par le truchement de mécanismes internes : chemins neuronaux ipsilatéraux et sous-corticaux se mettant en place ;

- Par le truchement de mécanismes externes : activités motrices qui distribuent ou transfèrent les informations à travers les hémisphères (mécanismes « étendus » d'intégration) :

    - Mouvements d'accès (mouvements de tête ou du corps) ;

    - Signaux croisés (expressions faciales et gestes accessibles par les deux hémisphères)

D'après Susan Hurley, il y a une analogie trop forte entre ces deux types d'intégration (interne ou externe), pour que l'on puisse douter que dans chaque cas, elles soutiennent une conscience unifiée. On pourrait se dire qu'il y a quand même une différence de taille car il semble naturel de supposer que l'on puisse avoir un accès conscient direct aux processus externes d'intégration et que, dans ce cas, il n'en résulterait pas une conscience unifiée. Mais l'argument a une portée limitée. Tout d'abord, qu'une conscience unifiée puisse s'enraciner sur des processus externes d'intégration n'implique pas nécessairement que ceux-ci sous-tendent toujours une conscience unifiée. Ensuite, on peut imaginer que, par un entraînement adéquat, introspection, méditation, biofeedback, etc. des sujets pratiquant une intégration interne puissent avoir une conscience directe de ces processus internes. Inversement, chez des sujets pratiquant pendant de longues années une intégration externe, celle-ci peut finir par devenir un automatisme auquel il ne prête plus attention, un peu comme lorsque l'on se familiarise avec la conduite automobile. Il ne semble donc pas que

l'on puisse véritablement arguer d'une différence dans la structure de la conscience sous-tendue par ces divers mécanismes intégratifs internes et externes, sur la base d'une différence dans l'appréhension consciente ou non de ces mécanismes eux-mêmes. On peut encore pousser plus loin l'analogie :

> « Supposons qu'un acallosal externalement intégré et avec une conscience unifiée soit placé dans des conditions expérimentales qui empêchent les mécanismes externes de fonctionner. En l'absence de mécanismes internes sur lesquels se rabattre, cela serait équivalent à une division chirurgicale quoique temporaire : cela altérerait la structure de la conscience. Par contraste, on pourrait soutenir que placer un patient ayant récemment subi une commissurotomie, dans ces conditions expérimentales, n'aurait pas la structure de sa conscience altérée. Cela bloquerait simplement certains canaux de communication entre deux centres séparés de conscience. » [230]

Nous pouvons conclure, avec Susan Hurley, que :

> « en principe, un mécanisme d'intégration partiellement externe pourrait être une partie du système causal qui supporte l'unité même de la conscience. Il n'y a rien de magique dans cette possibilité : elle en appelle à un système de causes et d'effets d'une manière parfaitement naturaliste, quand bien même certaines chaînes causales sont externes. En effet, cela dépend de la reconnaissance du fait qu'il n'y a aucune limite causale magique autour du cerveau qui empêcherait en principe les véhicules d'une conscience unifiée de s'étendre au-delà de lui. C'est l'intuition de l'*externalisme véhiculaire*. A la place, nous pouvons concevoir la base sub-personnelle d'une conscience unifiée comme une sorte de singularité dynamique dans le champ de flux causaux : un enchevêtrement de réactions bouclées d'orbites variées, certaines internes au cerveau ou au corps, d'autres impliquant partiellement des réactions motrices externes. Une telle singularité dynamique est centrée

[230] HURLEY, Susan. "Action, the unity of consciousness, and vehicle externalism", p. 9

> sur l'organisme et se déplace aux alentours avec lui, mais il n'a pas de limites nettes. » [231]

Le cas de Trevarthen est cependant problématique d'au moins 4 façons distinctes :

1) Qu'arrive-t-il à l'information du champ visuel droit qui n'est plus verbalement accessible ? Est-elle :

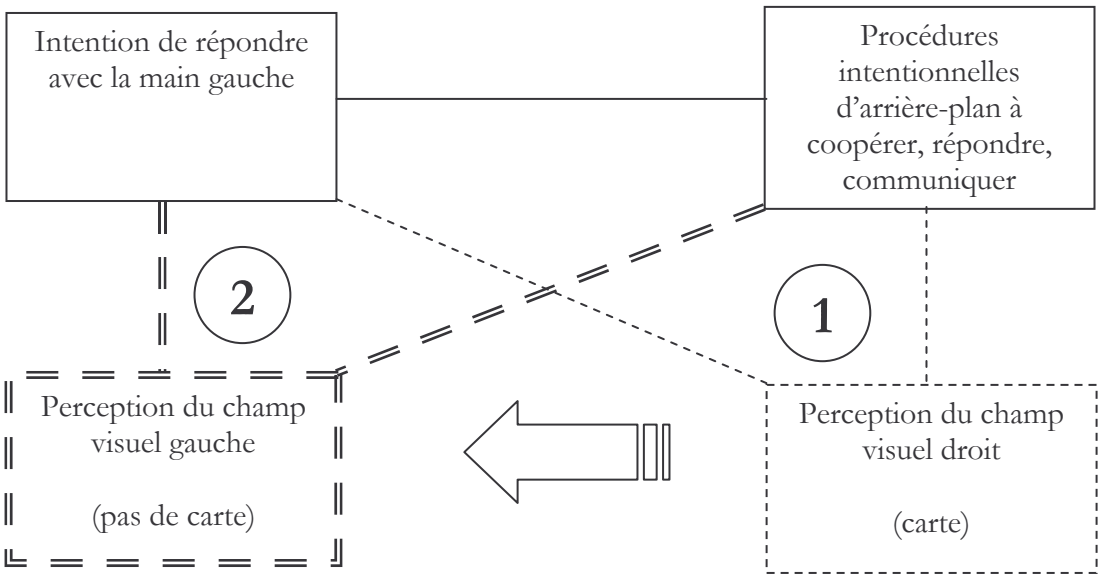
   a. Séparément consciente ?

   b. Plus consciente du tout ?

2) L'unité partielle de la conscience est-elle une alternative viable entre ces deux autres possibilités ?

3) Il se pourrait également que cela soit, non pas la structure de la conscience, mais son contenu qui dépendrait directement des intentions motrices (cf. schéma ci-dessous).

[231] HURLEY, Susan. "Action, the unity of consciousness, and vehicle externalism", p. 9




Selon ce modèle, il n'y aurait ni unité partielle ni centres de conscience séparés à aucun moment.

4) Enfin, le cas de Trevarthen pourrait être interprété en terme de changement d'attention, et non de changement d'intention, le second impliquant souvent le premier (cf. la théorie prémotrice de l'attention).

Toutefois, Susan Hurley répond à ces questions dans la suite de son article, en s'appuyant sur l'analyse d'une étude plus complexe, menée par Marcel[232], sur des sujets « normaux » cette fois-ci, mais en utilisant un protocole expérimental intéressant où des dissociations de la conscience sont rendues possibles grâce à des contraintes liées à des

[232] MARCEL, Anthony J. "Slippage in the Unity of Consciousness". *In:* BOCK, Gregory R. and MARSH, Joan (Eds.). *Experimental and Theoretical Studies of Consciousness.* Ainsi que : MARCEL, Anthony J. "What is Relevant to the Unity of Consciousness?". *In:* PEACOKE Christopher (Ed.). *Objectivity, Simulation and the Unity of Consciousness.*

types de réponses à fournir à un stimulus lumineux, aux modalités pour y répondre, et au temps imparti pour répondre.

Ce qu'il est important de souligner, c'est que dans cette étude, on voit apparaître **différentes intentions motrices simultanément**. Or, l'on observe malgré tout une dissociation entre ces intentions motrices. Par conséquent, on ne peut souscrire ici à une interprétation faisant l'économie de l'hypothèse d'une conscience partiellement unifiée, et ce pour des raisons d'intentions motrices, et non pas d'attention.

## 2. Les jumeaux « siamois »

### a) Les jumeaux « siamois » avec partage de tissu non-cérébral.

Parmi les anomalies d'origine embryonnaire, il en existe une, très rare (environ un cas répertorié sur 250 000 grossesses) mais très connue du grand public et particulièrement surprenante pour ne pas dire effrayante. Il s'agit des frères et sœurs « siamois » (en souvenir de Eng et Chang, originaires du Siam et qui furent deux frères siamois célèbres). Les frères et sœurs siamois sont des jumeaux issus d'un même ovule fécondé ou zygote (il s'agit donc de jumeaux monozygotes ou vrais jumeaux). Toutefois, chez les « siamois », la division de l'embryon se fait tardivement (au stade de bouton embryonnaire) et est incomplète. Les deux embryons résultants ont non seulement le chorion et le placenta en commun mais également plusieurs organes embryonnaires. On parle donc aussi de « jumeaux conjoints » pour les désigner (ce qui est un vocable plus approprié que nous utiliserons par la suite). Jusqu'à une époque récente, les « jumeaux conjoints » étaient des phénomènes que l'on

exhibait dans les foires. Actuellement, avec les progrès de la chirurgie, des séparations sont possibles dans bien des cas (mais elles ne sont pas toujours conseillées, lorsqu'elles impliqueraient qu'au moins un des deux jumeaux se trouve ainsi amputé de certains organes comme des membres par exemple). De plus, les échographies permettent souvent de détecter l'anomalie suffisamment tôt pour qu'une I.V.G. soit proposée.

Le cas des jumeaux conjoints pose problème à une conception corporelle de l'identité personnelle. En effet, si nous sommes notre corps, alors comment traiter le cas des jumeaux conjoints ? Toutes les études et même le bon sens indiquent clairement que les jumeaux conjoints sont deux personnes différentes, avec leur personnalité, leur caractère, leur conscience. Un des deux peut mourir sans que cela entraîne la mort de l'autre sur le coup (ce fut le cas avec Eng et Chang qui moururent avec quelques heures de décalage). Il y a indéniablement deux espaces phénoménologiques en présence et celui de l'un n'est pas numériquement identique à celui de l'autre.

Pour autant, les organes en commun peuvent difficilement être attribuables à un des deux jumeaux plutôt qu'à l'autre. S'il s'agit d'un organe moteur, deux volontés différentes peuvent le mouvoir. S'il s'agit d'un organe sensitif, ce qui se passe au niveau de cet organe est ressenti par deux sensibilités.

### b) Les jumeaux « siamois » avec partage de tissu cérébral.

Les jumeaux conjoints sont regroupés en divers types en fonction de leur point de jonction. Ainsi, les *craniopagus* sont des jumeaux conjoints au niveau du crâne (en général le

sommet). Cette anomalie est en générale viable et les jumeaux sont souvent séparables (au prix parfois de traumatismes cérébraux plus ou moins sévères) pour autant qu'ils ne partagent de tissu cérébral (ce qui est heureusement le cas le plus fréquent).

Le cas où les cerveaux partagent du tissu cérébral est le plus intéressant pour nous. Il offre un challenge aux conceptions de l'identité personnelle basées sur le cerveau ou la conscience. En effet, comment est la vie psychique de ces individus, de l'intérieur ? Sont-ils une conscience, deux, une et deux, une et demie (à supposer que cela veuille dire quelque chose de parler d'une conscience et demie !) ? Certes, on peut raisonnablement supposer qu'ils partagent certaines activités mentales. Néanmoins, tout porte à croire que si on détruit un des cerveaux, on tue quelqu'un et on laisse quelqu'un d'autre en vie. A partir de quel degré de fusion des cerveaux cette opération n'est plus un meurtre mais juste une lobotomie ? A partir de quel degré de fusion des cerveaux a-t-on affaire à une seule personne avec une seule conscience ? Hélas, les données cliniques et les études sur le sujet sont quasi-absentes, et nous en sommes donc réduit à la seule spéculation…

La théorie externaliste, jointe à une ontologie du processus, peut nous permettre de comprendre ce qui se passe. Comme pour les cas de patients split brain, nous sommes ici face à un cas de conscience partiellement unifiée. Les activités mentales qui s'enracinent dans les zones cérébrales partagées sont co-conscientes avec les activités mentales des deux cerveaux, bien que les activités mentales de ces deux cerveaux – outre la zone partagée – ne sont pas co-conscientes entre elles. Ainsi, c'est bien de deux consciences numériquement distinctes dont il est question, ces deux consciences ayant numériquement un contenu partiellement identique.

# Partie 3 : Identité personnelle et expériences de pensée

"A boy is about to go on his first date, and nervously considers what to talk about. He asks his father for advice. The father replies: 'My son, there are three subjects that always work. The famous three 'F's: 'Food', 'Family', and ... 'Filosophy'.' The boy picks up his date and they go to a soda fountain. Ice cream sodas in front of them, they stare at each other for a long time as the boy's nervousness builds. He remembers his father's advice, and chooses the first topic. He asks the girl: 'Do you like potato pancakes?' She says 'No', and the silence returns. After a few more uncomfortable minutes, the boy thinks again of his father's suggestion and turns to the second item on the list. He asks, 'Do you have a brother?' Once more, the girl says 'No' and there is silence again. The boy then plays his last card. He thinks of his father's advice and asks the girl: 'If you had a brother, would he like potato pancakes?'"

(Daniel Cohnitz, "Personal Identity and the Methodology of Imaginary Cases")

# I. Introduction

L'usage des expériences de pensée en général, et en particulier dans le domaine des réflexions sur l'identité personnelle, n'est pas sans poser problème. Dans cette dernière partie de notre thèse, nous aimerions en dire quelques mots. Nous pensons, en effet, que bien souvent ces critiques manquent leur cible. De plus, l'usage des expériences de pensée est si fréquent et central dans les discussions sur l'identité personnelle, que nous ne pouvons pas ignorer les critiques qui remettent complètement en cause la validité de cette stratégie, ou, du moins, remettent en cause sa validité lorsqu'elle s'applique au sujet qui nous occupe.

Il existe, bien sûr, des critiques portant sur certaines expériences de pensée en particulier. On peut les subdiviser en trois groupes :

1. Critiques d'incohérences : Dans certaines expériences de pensée, le scénario décrit serait en un certain sens, à une certaine profondeur d'analyse, incohérent, c'est-à-dire qu'il comporterait une contradiction interne.

2. Critiques des intuitions soi-disant évoquées par le scénario : Suite à la description d'une expérience de pensée, l'auteur clame généralement qu'intuitivement, nous serions conduits plutôt à dire ceci ou à dire cela, à

évaluer le scénario de telle manière ou de telle autre manière. On peut toutefois contester cette évaluation.

3. Critiques d'inapplicabilité : A supposer que le scénario décrit soit cohérent et que l'on tombe d'accord sur la manière de l'évaluer, on peut contester la généralisation de cette évaluation aux cas ordinaires, actuels.

Dans cette thèse, nos discussions ont surtout porté sur le point 1 et sur le point 2. Nous n'avons, pour ainsi dire, jamais considéré le point 3. Nous le ferons un peu plus loin, et nous approfondirons le point 2, à travers l'analyse de critiques, que l'on doit à la philosophe Tamar Szabò Gendler, de l'usage des expériences de pensée dans le cadre du problème de l'identité personnelle.

En neuropsychologie cognitive, la méthode consistant à observer un système cognitif altéré (paradigme lésionnel) pour en inférer le fonctionnement du système cognitif normal a fait ses preuves. L'anormal peut aider à comprendre le normal. Pourquoi en serait-il différent en philosophie ? Pourtant, concernant la méthodologie des expériences de pensée en général, un certain nombre d'auteurs, particulièrement des philosophes, se sont révélés sceptiques au sujet de la légitimité de l'inférence, à partir de cas au mieux exceptionnels, de conclusions au sujet des cas ordinaires ou généraux.

Kathleen Wilkes[233], par exemple, est très méfiante par rapport aux expériences de pensée de Parfit avec des gens se divisant comme des amibes. Elle déclare que nous ne savons tout simplement pas quoi dire quand nous pensons à cette sorte de chose. Pour elle,

---

233 WILKES, Kathleen V. *Real People: Personal Identity without Thought Experiments*.

une expérience de pensée ne devrait pas violer ce que nous considérons être les « lois de la nature ». Cela écarte les exemples de Parfit. Elle est proche, en cela, de Quine et Wittgenstein. Pour Quine, « Chercher ce qui est *logiquement requis* pour l'identité des personnes à partir de situations sans précédent, c'est suggérer que les mots ont une certaine force logique, au-delà de ce dont nos besoins passés les ont investis. »[234]. Quant à Wittgenstein, il déclare que : « Si vous imaginez certains faits autrement, les décrivez autrement que de la manière habituelle, alors vous ne pouvez plus imaginer l'application de certains concepts, parce que les règles pour leur application n'ont pas d'analogue dans la nouvelle situation »[235]. Dans le cas de Quine, l'idée semble être qu'une perturbation majeure de certaines parties centrales de notre réseau de croyances conduirait, à cause de l'holisme radical dont le philosophe se fait le défenseur, à une perturbation globale du sens. Cependant, un holisme si fort semble peu plausible.

En effet, si la signification d'un terme ou du contenu d'un concept était déterminée par l'ensemble du système de croyance dans lequel il s'insère, alors dans le monde actuel, cela aurait pour conséquence que pas deux personnes ne pourraient partager ne serait-ce qu'une seule croyance. De même, pas deux personnes ne pourraient entendre la même chose par ce qu'elles diraient. En effet, toute différence minime en un endroit de leur système cognitif entraînerait une différence du système entier. Cela semble assez peu vraisemblable.

Mais le caractère « fantaisiste » des expériences de pensée, n'est pas le seul aspect contesté de ces dernières.

---

234 QUINE, Willard Van Orman. "Review", 490.

235 WITTGENSTEIN, Ludwig. *Zettel*, § 350

A la suite des travaux des psychologues Daniel Kahneman et Amos Tversky[236], démontrant que dans un large nombre de cas, le raisonnement humain est entaché d'illusions cognitives, un certain nombre de philosophes[237] remettent en question la validité de l'intuition, et donc de la méthodologie de expériences de pensée. Dans ces dernières, il existe un risque consistant à accepter d'emblée, sans s'en rendre compte, un concept, une application d'un concept, une réalité, une possibilité, une nécessité ou un raisonnement, comme ne posant pas de problème, alors qu'il serait en fait très problématique voire douteux, après réflexion. Aux illusions perceptives, il y a donc, pourrait-on dire, un pendant cognitif, qui affecte le raisonnement humain dans un très grand nombre de cas. Pour bien faire comprendre le problème, on peut reprendre l'exemple de l'expérience de pensée de Frank Jackson[238] :

Mary est une personne qui, depuis sa naissance, a été élevée dans un univers entièrement en noir et blanc (pour être plus réaliste, on peut imaginer que ses yeux présentent une anomalie lui empêchant de distinguer autre chose que des nuances de gris). Elle a cependant pu apprendre (à travers des documents en noir et blanc) tout ce qui concernait la physique des couleurs, ainsi que tout ce qui concernait les processus perceptifs, cognitifs et cérébraux de traitement des couleurs. Son expertise en ce domaine est telle qu'en observant l'activité cérébrale d'un sujet en train de regarder un objet coloré (mettons une tomate), elle est capable de dire de quelle couleur ce dernier fait l'expérience (par exemple le rouge).

---

236 TVERSKY, Amos, et KAHNEMAN, Daniel. "The framing of decisions and the rationality of choice".

237 Cf. DEPAUL, Michael R. and RAMSEY, William. "Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry".

238 JACKSON, Frank. "Epiphenomenal Qualia"

Un jour, Mary est autorisée à sortir de son univers en noir et blanc (ou si l'on préfère : on parvient à corriger l'anomalie de ses yeux qui l'empêchait de percevoir les couleurs). A ce moment là, nous dit Jackson, il semble bien que Mary va apprendre quelque chose de nouveau concernant les couleurs : l'effet que cela fait des les expérimenter. En effet, voyant une tomate dans un jardin, il se peut qu'elle se dise : « Ah ! C'est donc à cela que cela ressemble le rouge ! ».

Or, par hypothèse, Mary connaissait déjà tous les faits physiques concernant les couleurs. Si elle apprend quelque chose de nouveau à leur sujet, cela ne peut donc pas être un nouveau fait physique. Conclusion : il existe des faits phénoménaux connaissables uniquement en en faisant l'expérience directe, et ces faits phénoménaux ne se réduisent pas à des faits physiques. Donc le matérialisme est faux.

On peut clarifier l'argument comme suit :

1. Avant de sortir de son univers en noir et blanc, Mary connaît déjà tous les faits physiques pertinents impliqués dans la perception des couleurs. (prémisse)

2. Avant de sortir de son univers en noir et blanc, Mary ne connaît pas l'effet que cela fait d'expérimenter les couleurs. (prémisse)

3. Après être sortie de son univers en noir et blanc, Mary connaît l'effet que cela fait d'expérimenter les couleurs. (prémisse)

4. Si un fait X est numériquement identique à un fait Y, alors connaître le fait X revient à connaître le fait Y, et inversement. (prémisse)

5. L'effet que cela fait d'expérimenter les couleurs n'est pas numériquement identique aux faits physiques pertinents impliqués dans la perception des couleurs. (de 1, 2 et 3)

6. Si l'effet que cela fait d'expérimenter les couleurs n'est pas numériquement identique aux faits physiques pertinents impliqués dans la perception des couleurs, alors le matérialisme est faux. (prémisse)

7. Le matérialisme est faux. (de 4 et 5)

Bien entendu, la logique du raisonnement est implacable, et les différentes prémisses semblent de prime abord de bon sens et acceptables. Toutefois, une analyse plus poussée peut nous en faire douter. Un certain nombre de présupposés implicites posent problème :

1. Est-il vraiment possible d'apprendre tout ce qui concerne la physique des couleurs, ainsi que tout ce qui concerne les processus perceptifs, cognitifs et cérébraux de traitement des couleurs ? A-t-on les capacités cognitives pour se faire ? Ce n'est pas sûr. Certains concepts nécessaires à la compréhension physicaliste des couleurs pourraient nous être inaccessibles, notre puissance de calcul pourrait être trop limitée, notre mémoire insuffisante, etc.

2. Si, malgré tout, une connaissance absolue des aspects physiques des couleurs et de la vision des couleurs était humainement envisageable, serait-il vraiment impossible que cette connaissance soit équivalente à la connaissance des aspects phénoménaux de la couleur ? Dennett imagine par exemple une deuxième Mary qui, en découvrant les couleurs, s'exclame : « Ah, la perception des couleurs est exactement comme je pensais qu'elle devrait l'être. ». N'est-ce pas tout aussi plausible ? Bien entendu, aucun neuroscientifique, à l'heure actuelle, ne serait en mesure de retrouver l'impression perceptive des couleurs à partir des connaissances physiques et physiologiques les plus pointues dont on dispose en ce domaine (sinon les neuroscientifiques étudiant la perception chez les oiseaux seraient en mesure de se représenter en première personne ce que cela fait de voir le monde en quadrichromie, étant donné que c'est ainsi que, semble-t-il, les oiseaux voient le monde). Mais nos connaissances physiques et physiologiques *actuelles* des couleurs sont-elles tout ce qu'il y a à savoir en ce domaine ? On peut en douter.

3. Si durant une si longue partie de sa vie, Mary n'a pas expérimenté les couleurs, cela n'a-t-il aucune répercussion sur la capacité de son cerveau à les percevoir par après ? En sera-t-elle vraiment capable ? Si elle en est capable, comment être sûr qu'elle expérimentera les couleurs comme nous ? D'ailleurs, si les *qualia* sont des entités irréductiblement privées, comment est-il possible de parler de façon « adéquate », « anormale », « habituelle », etc. de les expérimenter ? Comment comparer ce qui est irréductiblement privé ?

4. La prémisse 4 est également contestable. On peut trouver, semble-t-il, un contre-exemple : Jean sait que l'aspirine soigne la migraine. Jean ignore que l'acide acétyle salicylique soigne la migraine. Pourtant, l'aspirine est numériquement identique à l'acide acétyle salicylique. Bien sûr, on peut répondre à ce contre exemple que le rapport épistémique entre un sujet et ses états phénoménaux n'est pas le même qu'entre ce sujet et des faits physiques. Ainsi, on pourrait soutenir que la connaissance des faits physiques est contingente, tandis que la connaissance de nos propres états phénoménaux serait nécessaire. Mais la critique wittgensteinienne du langage privé et l'expérience de pensée de Dennett avec les cafetiers Chase et Sanborn, tendent à montrer qu'il n'y a pas de sens à parler de connaissance de nos états phénoménaux si cette connaissance est tenue pour infaillible et privée, et que sinon, la connaissance de nos états phénoménaux se trouve être contingente comme celle de n'importe quel fait physique.

5. Même si l'effet que cela fait d'expérimenter les couleurs n'est pas numériquement identique aux faits physiques pertinents impliqués dans la perception des couleurs, cela ne signifie pas nécessairement que le matérialisme est faux. Ce qu'apprendrait Mary lorsqu'elle découvre le monde des couleurs pourrait ne pas être de nature factuelle. Cela pourrait être une nouvelle manière de concevoir des faits déjà connus, une nouvelle aptitude cognitive, etc. Il semble de toute manière curieux que ce qui serait, par hypothèse, ineffable, privé, intrinsèque, incorrigible, etc. puisse être catégorisé comme fait. En général, on entend par fait, ce qui

est objectivement constatable, communicable, dont la connaissance – faillible – dépend de certains principes épistémologiques, méthodologiques, théoriques, etc., et dont la nature est relationnelle et non pas intrinsèque.

A l'issu de l'analyse de cet exemple, et sans vouloir prétendre à l'exhaustivité, on peut répertorier un certain nombre d'illusions cognitives inhérentes à au moins certaines expériences de pensée (celle-ci entres autres) :

1. Une surestimation des possibilités cognitives, biologiques, physiques, métaphysiques dans certains cas, car est-il vraiment possible de connaître absolument tout ce qui concerne les faits physiques concernant les couleurs ?

2. Une sous-estimation des possibilités cognitives, biologiques, physiques, métaphysiques dans d'autres cas. Car s'il est possible de tout connaître des faits physiques concernant les couleurs, qu'est-ce qui nous assure que nous ne connaîtrions pas du même coup les aspects phénoménaux de leur perception ? Et n'est-il pas possible que ce qui soit physiquement pertinent dans la perception des couleurs excède ce que l'on connaît actuellement et ce que l'on peut connaître ?

3. Toujours une sous-estimation des possibilités ou peut-être un manque d'informations empiriques. En effet, le cerveau de Mary peut avoir été profondément marqué par l'absence, durant tout ce temps, de

stimulations colorées. On a aussi des confusions conceptuelles, ce que Ryle appelle des *erreurs de catégorie*[239]. Car, comment Mary peut-elle savoir que ce qu'elle expérimente en regardant une tomate est de même nature que ce que le sujet dont elle examine le cerveau expérimente en voyant une tomate si les qualia sont irréductiblement subjectifs ?

4. Toujours des confusions conceptuelles ou alors un manque de considération de certains contre-exemples (avec l'aspirine et l'acide acétyle salicylique).

5. Une surestimation de la portée de certains résultats, de certaines conclusions. Cela peut être lié à des confusions conceptuelles qui nous font attribuer à certains concepts un poids ontologique trop élevé (Mary n'apprend pas nécessairement de nouveaux faits, mais possiblement de nouvelles façons d'appréhender les mêmes faits ou de nouvelles habiletés), ou à une mauvaise compréhension des théories.

De prime abord, il semble que l'on peut dégager trois grands groupes d'illusions cognitives :

1. **Illusions modales :** On évalue mal les possibilités, impossibilités, nécessités, de telle ou telle proposition, au regard de considérations cognitives, biologiques, physiques, métaphysiques, etc.

---

239 RYLE, Gilbert, *The Concept of Mind.*

2. **Illusions conceptuelles :** On croit, à tort, que le concept que l'on emploie à un sens/référent – qui plus est un sens/référent unique et univoque – dans le contexte dans lequel on l'emploie, et tout au long de notre raisonnement.

3. **Illusions épistémiques :** On croit à tort disposer de toutes les connaissances pertinentes pour réfléchir à un problème donné. On ignore ou connaît mal ou néglige comme sans importance, un certain nombre de faits qui peuvent avoir leur importance. Parmi ces faits, une présentation – volontairement ou involontairement – caricaturale d'une théorie que l'on veut réfuter peut donner à croire faussement qu'on a réussi à la réfuter.

Il semble pourtant exagéré de tirer de ces considérations une condamnation sans appel de toute la méthodologie des expériences de pensée. Il suffit de remarquer que, pour la plupart, ces illusions cognitives se rencontrent dans d'autres activités humaines, par exemple en sciences. Ce n'est pas une raison pour disqualifier la démarche scientifique en tant que telle. Ce que l'existence de ces illusions cognitives nous enseigne, ce n'est pas l'invalidité pure et simple de la méthodologie des expériences de pensée, mais à être prudent et rigoureux dans leur mise en place.

Toutefois, l'on pourrait arguer que même un philosophe s'évertuant à être le plus prudent et rigoureux dans ses expériences de pensée ne pourrait pas aboutir à une conclusion valable. Souvent, par les expériences de pensée, on cherche, semble-t-il, à atteindre une certaine vérité métaphysique. Même s'il prenait soin d'analyser correctement

ce qui est cognitivement, biologiquement, et physiquement possible, nécessaire, impossible, comment notre philosophe pourrait-il évaluer une possibilité, nécessité ou impossibilité proprement métaphysique ? *A fortiori*, même s'il prenait soin de se renseigner dûment sur les faits cognitifs, biologiques, physiques, etc., comment pourrait-il acquérir une connaissance de faits métaphysiques ?

Traditionnellement, on considère que la concevabilité est un bon guide de la possibilité métaphysique. Tout ce qui serait concevable serait métaphysiquement possible. Mais sur quoi repose un tel postulat ? Est-il fondé ? N'existe-t-il pas des choses dont on peut montrer qu'elles sont concevables quoique métaphysiquement impossibles ?

Dans le chapitre suivant, nous analyserons un article de David J. Chalmers[240] qui propose à la fois une réponse à cette objection, et un certain nombre d'outils conceptuels permettant d'éviter certains écueils dans le passage de la concevabilité à la possibilité métaphysique.

Puis, nous analyserons, de manière détaillée, deux articles de Tamar Gendler, qui concentrent leur critique sur l'usage des expériences de pensée dans le cadre des réflexions sur l'identité personnelle.

---

[240] CHALMERS, David John. "Does Conceivability Entails Possibility?". *In:* GENDLER, Tamar Szabò and HAWTHORNE, John. *Conceivability and Possibility*

## II. Du concevable au possible

Les argumentations *a priori* sur le possible et le nécessaire d'un point de vue métaphysique[241], prennent souvent cette forme :

1. On part d'une proposition de nature épistémique (sur ce qui peut être connu ou conçu)

2. On en infère une proposition de nature modale (sur ce qui est possible ou nécessaire)

3. Puis on passe de cette proposition de nature modale à une proposition de nature métaphysique (sur la nature des choses du monde)

Mais comment relier les domaines épistémiques et modaux ? Plus particulièrement : comment relier la concevabilité à la possibilité ? Le passage du concevable au métaphysiquement possible est critiqué par de nombreux philosophes. Il existerait, entre autres, de nombreux contre-exemples :

- Certaines propositions mathématiques fausses (telle que la conjecture de Goldbach ou bien sa négation) seraient concevables (on pourrait

[241] Il convient peut-être de distinguer entre possibilité physique/naturelle et possibilité métaphysique. Ainsi, une vitesse supraluminique semble concevable, mais elle est physiquement et naturellement impossible car elle contredit les lois de la physique et de la nature. Cependant, une telle vitesse pourrait être métaphysiquement possible, puisqu'il pourrait bien exister des mondes métaphysiquement possibles avec des lois physiques et naturelles différentes.

s'imaginer un comité d'experts en mathématiques annonçant qu'on l'a démontrée ou réfutée) mais impossibles ;

- Les identités *a posteriori* : il serait concevable mais impossible que Hesperus ne soit pas Phosphorus, et que l'eau ne soit pas $H_2O$.

Face à ces objections, Chalmers nous suggère de distinguer plusieurs sens à « concevable ».

## *A. Différentes manières de concevoir la concevabilité*

Le philosophe australien, dans son article, procède à une analyse de trois paires de distinctions principales. Celles entre :

1. Concevabilité Prima faciès vs Idéale

2. Concevabilité Négative vs Positive

3. Concevabilité Primaire vs Secondaire

Sur presque 10 pages, David Chalmers s'emploie à décortiquer ces différentes manières de concevoir la concevabilité. Nous nous permettrons de résumer de manière synthétique ses distinctions à travers le tableau suivant :

Les deux premières distinctions sont combinables avec la troisième :

| S est concevable : | **Primairement/ Epistémiquement :** On part des situations concrètes du monde tel qu'il est et qui ont donné leur sens aux mots et expressions ou ont enrichi celui-ci, et on se demande si S est concevable (dans un monde où les événements se seraient déroulés autrement, où les choses auraient été différentes).<br><br>*Par ex. : l'eau dans les mers aurait pu ne pas avoir pour formule $H_2O$ mais, par ex., XYZ.* | **Secondairement/ Subjonctivement :** On part des définitions actuelles, du sens et de la référence qu'ont les mots et expressions dans le monde tel qu'il est, et on se demande si, ceci étant fixé, S est concevable (dans un monde où les événements se seraient déroulés autrement, où les choses auraient été différentes).<br><br>*Par ex. : l'eau est nécessairement une substance de formule $H_2O$, toutefois, les mers auraient pu contenir un liquide aqueux de formule XYZ, mais ce ne serait pas de l'eau.* |
|---|---|---|
| **Prima Faciès :**<br>si S est concevable de prime abord, sur les premières apparences, superficiellement.<br><br>*Par ex. : la fausseté du dernier théorème de Fermat est concevable de prime abord, en imaginant un comité d'experts en mathématiques annonçant qu'il a été démontré faux.* | *Par ex. : la fausseté du dernier théorème de Fermat est concevable de prime abord, en imaginant un comité d'experts en mathématiques annonçant qu'il a été démontré faux.* | *Par ex. : Quelqu'un ne s'y connaissant pas en chimie pourrait concevoir secondairement la proposition : « L'eau n'est pas identique à $H_2O$. »* |
| **Idéalement :**<br>si S est concevable sur la base d'une réflexion rationnelle idéale – un raisonneur idéal, sans limitation cognitive, peut concevoir S – ou encore : lorsqu'il y a un sujet pour lequel S est concevable prima faciès, appuyée par une justification qui ne peut pas être démolie par un meilleur raisonnement.<br><br>*Par ex. : une réflexion rationnelle réfute la concevabilité prima faciès de la fausseté du dernier théorème de Fermat.* | *Par ex. : une réflexion rationnelle réfute la concevabilité prima faciès de la fausseté du dernier théorème de Fermat, mais pas celle d'une formule chimique de l'eau qui aurait été XYZ au lieu de $H_2O$.* | *Par ex. : l'apport de connaissances en chimie montre qu'on ne peut concevoir secondairement la proposition : « L'eau n'est pas identique à $H_2O$ », mais cela n'interdit pas de concevoir que les mers auraient pu contenir autre chose que de l'eau, par exemple un liquide aqueux de formule XYZ.* |

| | | |
|---|---|---|
| **Négativement :**<br>lorsque S ne peut pas être exclu ou écarté a priori par nos connaissances ou nos croyances au sujet du réel ou lorsqu'il n'y a (apparemment) pas de contradiction en S.<br><br>*Par ex. : « Il existe des choses inconcevables » est une proposition qui semble ne pas pouvoir être écartée a priori. Elle ne semble pas receler de contradiction, elle n'apparaît pas en contradiction avec nos connaissances et croyances au sujet du monde.* | *Par ex. : une idée verte qui dort furieusement est négativement, primairement, concevable.* | *Par ex. : une idée verte qui dort furieusement est négativement, primairement, inconcevable (on ne peut pas appliquer au concept d'idée un prédicat coloré).* |
| **Positivement** :<br>lorsque l'on peut former une conception positive d'une situation dans laquelle S est le cas, i.e. lorsque l'on peut imaginer que S, i.e. lorsque l'on peut imaginer une situation qui vérifie S.<br><br>*Par ex. : On ne peut pas imaginer une situation qui vérifie : « Il existe des choses inconcevables ».*<br><br><u>Distinctions :</u><br><br>Imagination<br>↳ … perceptive \| *Par ex., S1 : « les cochons volent. », vérifié par le fait de s'imaginer perceptivement des cochons en train de voler.*<br>↳ … modale \| *Par ex., S2 : « L'Allemagne a gagné la Seconde Guerre Mondiale », vérifié par le fait de s'imaginer l'armée allemande gagner certaines batailles et terrasser les forces alliées à travers l'Europe.*<br>↳ … cohérente : Une situation est modalement imaginable avec cohérence s'il est possible d'en remplir n'importe quel détail de telle manière qu'aucune contradiction n'apparaisse. \| *Par ex., il semble que l'on puisse s'imaginer modalement et positivement une situation qui vérifie la proposition : « Des voyages dans le passé sont métaphysiquement possibles ». Toutefois, si l'on raisonne sur une telle situation, des contradictions apparaissent (par ex. le paradoxe du grand-père), si bien que cette situation ne peut plus raisonnablement être tenue pour telle qu'elle vérifie cette proposition.* | *Par ex. : une idée verte qui dort furieusement est positivement, primairement, inconcevable (on ne peut pas s'imaginer une situation qui la vérifie).* | *Par ex. : une idée verte qui dort furieusement est positivement, secondairement, inconcevable (on ne peut pas s'imaginer une situation qui la vérifié et on ne peut de toute façon pas appliquer au concept d'idée un prédicat coloré)* |

Les distinctions Prima Faciès Vs. Idéalement et Négativement Vs. Positivement, sont également combinables entre elles…

| *Par ex. : S3 : « Il existe des créatures qui voient le monde en quadrichromie. », et S4 : « Il existe des mammifères sans vertèbres. ».* | | | |
|---|---|---|---|
| **Prima Faciès** | **Négativement** | S3 et S4 peuvent être, prima faciès, négativement, primairement, concevables. | S3 et S4 peuvent être, prima faciès, négativement, secondairement, concevables. |
| | **Positivement** | S3 n'est pas, prima faciès, positivement, primairement, concevable (car les êtres humains ne peuvent percevoir qu'en trichromie). Mais S4 peut être, prima faciès, positivement, primairement, concevable. | S3 n'est pas, prima faciès, positivement, secondairement, concevable (car les êtres humains ne peuvent percevoir qu'en trichromie). Mais S4 peut être, prima faciès, secondairement, primairement, concevable. |
| **Idéalement** | **Négativement** | S3 et S4 sont idéalement, positivement/ négativement, primairement, concevables. | S3 est idéalement, positivement/ négativement, secondairement, concevable. S4 est idéalement, positivement/ négativement, secondairement, inconcevable (des connaissances en systématique nous enseignent que la classe des mammifères appartient à l'embranchement des vertébrés). |
| | **Positivement** | | |

Au bout du compte, il existe, a priori, 8 manières dont on peut dire concevoir une certaine proposition S.

### *B. Fossés entre la concevabilité et la possibilité :*

Pour Chalmers :

1. **La concevabilité Prima Faciès (surtout Négative) est un guide imparfait de la possibilité.** Par ex., la concevabilité Prima Faciès de la véracité ou de la fausseté de la conjecture de Goldbach, de figures géométriques impossibles, ou encore le paradoxe de la Faucheuse (une variante du paradoxe de Zénon).

2. **La concevabilité Positive est un meilleur guide de la possibilité que la concevabilité Négative.** Mais idéalement, les deux types de concevabilité tendent à se rejoindre.

3. **La concevabilité Primaire est un guide imparfait de la possibilité Secondaire.** On considère souvent qu'il est concevable que l'eau ne soit pas identique à $H_2O$ mais que dans les faits, cela est impossible. L'identité entre l'eau et $H_2O$, bien qu'a posteriori, est nécessaire (pour Kripke, si A est identique à B, alors *il est nécessaire que* A soit identique à B). En réalité, un tel raisonnement confond la concevabilité Primaire et la possibilité Secondaire. Si la concevabilité Primaire est un bon guide de la possibilité, c'est uniquement lorsqu'il s'agit de la possibilité Primaire : si, lorsque la structure chimique de l'eau a été identifiée, on avait découvert qu'elle avait pour formule XYZ, alors l'eau serait identique à XYZ. La possibilité Secondaire, elle, peut être bien cernée par la concevabilité secondaire : Puisque, lorsque l'on a identifié la structure chimique de l'eau, on n'a pas découvert qu'elle avait pour formule XYZ mais $H_2O$, alors l'eau est identique – *a posteriori*, mais nécessairement – à $H_2O$. Par conséquent, cette

donnée, rajoutée à la définition de l'eau, rend impossible de concevoir, Secondairement et Idéalement, de l'eau qui ne serait pas $H_2O$. A partir de là, si nous découvrions, sur une planète, une substance non aqueuse, dont la formule chimique serait $H_2O$, nous serions fondés à dire que c'est une forme d'eau particulière. Inversement, si une substance aqueuse de formule XYZ venait à être découverte, nous pourrions sans doute dire qu'elle *ressemble à* de l'eau, mais nous ne pourrions pas dire que c'est de l'eau. Egalement, si l'eau de nos océans venait à être remplacée intégralement par une substance aux mêmes propriétés, mais de formule XYZ, alors il n'y aurait à proprement parler plus d'eau dans les océans, mais un liquide équivalent de formule XYZ.

### *C. De la concevabilité à la possibilité : l'argument « pragmatique » de Chalmers*

Pour Chalmers :

1. La concevabilité Idéale, Négative/Positive, Primaire, implique la possibilité Primaire ;

2. La concevabilité Idéale, Négative/Positive, Secondaire, implique la possibilité Secondaire.

Dans la suite de son article (p. 20 à 30), Chalmers s'emploie à analyser un possible fossé entre la concevabilité Idéale Négative et la concevabilité Idéale Positive. Nous ne rentrerons pas dans les détails de cette analyse qui dépasserait le cadre de notre sujet, et

renvoyons le lecteur intéressé à l'article de Chalmers. Ce qui en ressort est que, à part pour quelques cas de propositions excentriques du genre « il existe des choses inconcevables », les concevabilités Négative et Positive convergent Idéalement. En pratique, nous pouvons les considérer comme équivalentes.

Pour notre propos, il est plus intéressant de se demander si l'on peut justifier d'une quelconque manière un passage de la concevabilité à la possibilité métaphysique. C'est ce qu'essaye de faire Chalmers à la fin de son article (p. 30 à 34).

Ce passage serait impossible s'il existait (réellement ou hypothétiquement) des « nécessités fortes » (*strong necessities*), c'est-à-dire des propositions dont la négation serait Primairement, Positivement – voire aussi Idéalement – concevable, mais qui seraient néanmoins vraies dans tous les mondes possibles (Primairement).

Pour que de telles nécessités existent, l'espace des situations concevables devrait dépasser celui des mondes possibles.

Chalmers affirme qu'il n'existe pas de clairs exemples de nécessités fortes et que les seuls candidats sont hautement tendancieux. Il s'emploie ensuite à examiner et à réfuter ces candidats potentiels que sont :

- L'existence de Dieu (conçu comme un être existant nécessairement) ;

- Les lois de la nature (parfois vues comme métaphysiquement nécessaires)

- Des concepts dont l'extension et le sens seraient fixés par nos réponses (proposition faite par Yablo[242])

- Les lois psychophysiques (pour certains matérialistes [de type B, selon la terminologie de Chalmers], les p-zombies sont Idéalement concevables mais métaphysiquement impossibles).

Nous ne rentrerons pas dans le détail de l'argumentation de Chalmers pour ces différents cas et renvoyons encore une fois le lecteur intéressé à l'article original.

Il apparaît toutefois que :

- Les contre-arguments apportés par Chalmers sont pertinents mais pas toujours décisifs : on peut toujours admettre, ne serait-ce qu'hypothétiquement, l'existence de nécessités fortes.

- La critique de potentiels contre-exemples à la validité du passage de la concevabilité à la possibilité métaphysique ne constitue pas une défense positive de la validité de ce passage. Il faut autre chose.

Autre chose, c'est ce qu'essaye d'apporter Chalmers au final, à travers un argument positif en faveur de la validité du passage du concevable au possible métaphysique. Cet argument repose sur la notion centrale de modalité. Rappelons la structure typique d'un argument *a priori* en faveur d'une possibilité/nécessité métaphysique :

---

[242] YABLO, Stephen. "Coulda, Woulda, Shoulda". *In:* GENDLER, Tamar Szabò and HAWTHORNE, John. *Conceivability and Possibility*

1. Proposition de nature épistémique (sur ce qui peut être connu ou conçu)

2. Proposition de nature modale (sur ce qui est possible ou nécessaire)

3. Proposition de nature métaphysique (sur la nature des choses du monde)

L'argument implique de localiser les racines de nos concepts modaux dans le domaine rationnel (*rationalisme modal*). Lorsque l'on examine la finalité dans laquelle la modalité s'inscrit, il est frappant de constater que l'essentiel de cette finalité est fermement lié au domaine du rationnel et du psychologique : analyser le contenu des pensées et la sémantique du langage, rendre compte de la pensée contrefactuelle, analyser l'inférence rationnelle, etc. Pour être utilisables en pratique dans de telles situations, les notions de possibilité et de nécessité doivent être des concepts modaux rationnels, liés constitutivement aux notions de consistance, d'inférence rationnelle ou de concevabilité. Il n'est pas difficile de montrer que, même si tous les mondes concevables ne sont pas des mondes métaphysiquement possibles, nous avons *besoin* de concepts modaux rationnels pour les analyser au mieux. Nous pouvons ici parler de possibilité *logique*. Par exemple, même si tous les mondes avec des lois de la nature différentes du nôtre sont métaphysiquement impossibles, il demeure extrêmement pratique d'utiliser un plus large espace des mondes (ou des entités « mondoïdes » [on peut les appeler, par exemple, des scénarios]) logiquement possibles avec des lois différentes du nôtre. En effet, cela nous aide à comprendre et à analyser l'activité d'un scientifique investiguant les lois de la nature, les hypothèses qu'il pose, les inférences qu'il pratique. Ainsi, un scientifique, dans son laboratoire, va se construire des modèles rationnels de l'agencement des faits observables afin de pouvoir effectuer des prédictions et des vérifications. Ce faisant, il émettra des

affirmations conditionnelles et s'engagera dans une pensée contrefactuelle au sujet de ces possibilités (du genre : « Si l'hypothèse H1 est juste, alors j'observerai le résultat expérimental E1 dans les conditions C1. »).

Cet espace logique de mondes possibles semble être tout ce dont nous avons besoin pour rendre compte de tous les phénomènes modaux en lesquels nous avons des raisons de croire. Si cet espace modal logique est tenu pour cerner tout ce qui est métaphysiquement possible ou nécessaire, alors nous sommes en présence d'un *monisme modal*, fondé sur un seul principe modal primitif (la possibilité logique). Le croyant en l'existence de nécessités fortes, par contraste, doit embrasser un *dualisme modal*, avec deux principes modaux primitifs distincts (la possibilité logique d'une part, la possibilité métaphysique d'autre part, irréductible l'une à l'autre). Il n'y a pas de raison d'accepter un tel dualisme modal, alors que le monisme modal suffit à expliquer tous les phénomènes non tendancieux. Il n'y a pas de données modales supplémentaires pour expliquer une modalité métaphysique distincte : ce qui a besoin d'être expliqué est déjà expliqué. Ce n'est pas juste un argument par la simplicité (par l'entremise du rasoir d'Occam). En effet, il n'y a tout simplement pas de concept distinct de possibilité métaphysique, à laquelle la seconde sorte de modalité serait une réponse potentielle. L'impression momentanée de l'existence d'un tel concept vient d'une compréhension confuse de distinctions ontiques/épistémiques telles que celles entre apriorité et nécessité, ou entre concept et propriété, chacune de ces distinctions pouvant aisément s'inscrire au sein d'un monisme modal avec l'aide d'une sémantique bidimensionnelle (Primaire/Epistémique Vs. Secondaire/Subjonctif).

Ultimement, il y a juste un cercle de concepts modaux, incluant à la fois les concepts modaux rationnels (validité, implication rationnelle, apriorité, concevabilité) et les concepts modaux métaphysiques (possibilité, nécessité, propriété).

## *D. De la concevabilité à la possibilité : un argument rationaliste*

L'argument de Chalmers est, on le voit, plutôt pragmatique et soucieux d'économie ontologique.

Nous pensons qu'un autre argument pourrait être plus convaincant, pour des personnes d'un esprit plus rationaliste que pragmatique. Voici l'argument que nous proposerions, pour notre part, à l'appui de la thèse suivant laquelle la concevabilité (Idéale) implique la possibilité métaphysique.

1. Tout sujet cognitif fait partie du réel métaphysique. (prémisse)

2. Si tout sujet cognitif fait partie du réel métaphysique, alors ce qui est nécessairement vrai du réel métaphysique est nécessairement vrai de tout sujet cognitif. (prémisse)[243]

3. Si ce qui est nécessairement vrai du réel métaphysique est nécessairement vrai de tout sujet cognitif, alors si une proposition p décrit une situation métaphysiquement impossible, alors la proposition p est cognitivement impossible. (prémisse)

4. Si ce qui est nécessairement vrai du réel métaphysique est nécessairement vrai de tout sujet cognitif, alors si une proposition p décrit une situation métaphysiquement nécessaire, alors la proposition p est cognitivement nécessaire. (prémisse)

5. Si une proposition p est cognitivement impossible, alors elle est inconcevable Idéalement. (prémisse)

---

[243] Car en effet, si A fait partie de B, alors ce qui est nécessairement vrai de B est nécessairement vrai de A. *Ex. : si les chats sont des mammifères et s'il est nécessairement vrai que les mammifères ont des vertèbres, alors il est nécessairement vrai que les chats ont des vertèbres.* La réciproque n'est pas vraie. *Ex. : si les chats sont des mammifères et s'il est nécessairement vrai que les chats miaulent, cela n'implique pas qu'il soit nécessairement vrai que les mammifères miaulent (par ex., les chiens ne miaulent pas).* (Tout cela ne vaut cependant pas pour ce qui est de la possibilité : si A fait partie de B, alors ce qui est possiblement – mais pas nécessairement – vrai de B n'est pas nécessairement possiblement vrai de A. *Ex. : si les chats sont des mammifères et s'il est possiblement vrai que les mammifères volent – par ex. les chauves-souris – cela n'implique pas qu'il soit possible que les chats volent.* En revanche, la réciproque est vraie. *Ex. : si les chats sont des mammifères et s'il est possiblement vrai que les chats sont gris, alors il est possiblement vrai que les mammifères sont gris [ex. : les chats peuvent l'être]*).

6. Si une proposition p est cognitivement nécessaire, alors elle est nécessairement concevable Idéalement. (prémisse)

7. Ce qui est nécessairement vrai du réel métaphysique est nécessairement vrai de tout sujet cognitif. (de 1 et 2)

8. Si une proposition p décrit une situation métaphysiquement impossible, alors la proposition p est cognitivement impossible. (de 7 et 3)

9. Si une proposition p décrit une situation métaphysiquement nécessaire, alors la proposition p est cognitivement nécessaire. (de 7 et 4)

10. **Si une proposition p décrit une situation métaphysiquement impossible, alors la proposition p est inconcevable Idéalement.**[244] **(de 8 et 5).** *Ex : 1/ Les cercles carrés sont métaphysiquement impossibles. Donc, Idéalement, on ne peut pas concevoir des cercles qui seraient des carrés. (C'est en effet le cas). 2/ Que la Vertu ait une couleur est métaphysiquement impossible. Donc, Idéalement, on ne peut pas concevoir que la Vertu serait rouge, verte, bleue, etc. (C'est en effet le cas).* | Formellement : $MI(p) \rightarrow II(p)$

[244] Notons que la réciproque n'est pas vraie. Une proposition qui serait Idéalement inconcevable pourrait néanmoins correspondre à une possibilité métaphysique (contingente toutefois). Son inconcevabilité Idéale pourrait correspondre simplement à une limitation de nos capacités cognitives. Ainsi, même si nous ne parvenons pas à nous concevoir comme inexistants, cela ne voudrait pas nécessairement vouloir dire que notre inexistence est métaphysiquement impossible. De même, l'impossibilité dans laquelle nous sommes de nous représenter un Dieu Un qui serait en même temps Trois Personnes n'est pas un argument décisif à l'encontre de son existence. Bien sûr, ces impossibilités cognitives pourraient très bien signifier tout simplement que notre concevabilité n'est pas Idéale dans ces cas-là. En effet, mais cela va dans le sens de ce que nous voulons dire : les impossibilités cognitives peuvent très bien excéder les impossibilités métaphysiques.

11. **Si une proposition p décrit une situation métaphysiquement nécessaire, alors la proposition p est nécessairement concevable Idéalement.[245] (de 9 et 6).** *Ex. : 1/ Il est métaphysiquement nécessaire que la somme des angles d'un triangle inscrit dans un plan euclidien fasse 180°. Donc, Idéalement, on ne peut pas faire autrement que de concevoir des plans euclidiens où les triangles ont la somme de leurs angles égale à 180°. (C'est en effet le cas). 2/ Il est métaphysiquement nécessaire que tout phénomène s'inscrive dans le temps. Donc, Idéalement, on ne peut pas concevoir un phénomène qui serait hors de tout temps. (C'est en effet le cas).* | Formellement : $MN(p) \rightarrow NCI(p)$

12. Une proposition p est métaphysiquement contingente, si et seulement si p est métaphysiquement possible mais n'est pas métaphysiquement nécessaire. (prémisse). | Formellement : $MC(p) \leftrightarrow \neg(MI(p) \vee MN(p)) \leftrightarrow \neg MI(p) \wedge \neg MN(p)$

13. Une proposition p est contingentement concevable Idéalement, si et seulement si p est concevable Idéalement mais n'est pas nécessairement concevable Idéalement. (prémisse). | Formellement : $CCI(p) \leftrightarrow \neg(II(p) \vee NCI(p)) \leftrightarrow \neg II(p) \wedge \neg NCI(p)$

[245] Notons que la réciproque n'est pas vraie. Une proposition qui serait Idéalement concevable nécessairement pourrait néanmoins correspondre à une possibilité métaphysique contingente (non nécessaire). La nécessité Idéale que nous concevons pourrait correspondre simplement à une caractéristique essentielle de nos capacités cognitives. Ainsi, même si nous nous concevons Idéalement comme éternels, cela ne voudrait pas nécessairement vouloir dire qu'il est métaphysiquement nécessaire que nous le soyons. De même, la nécessité Idéale que nous percevons dans les vérités logiques n'est pas un argument décisif en faveur de leur nécessité métaphysique. Bien sûr, ces nécessités cognitives pourraient très bien signifier tout simplement que notre concevabilité n'est pas Idéale dans ces cas-là. En effet, mais cela va dans le sens de ce que nous voulons dire : les nécessités cognitives peuvent très bien excéder les nécessités métaphysiques.

**14. Si une proposition p est contingentement concevable Idéalement, alors la proposition p est métaphysiquement contingente.[246] (de 10, 11, 12 et 13).**

*Ex. : 1/ Une mouche avec deux paires d'ailes est contingentement concevable Idéalement. Donc, une mouche avec deux paires d'ailes est métaphysiquement contingente (cf. les expériences sur les gènes homothétiques chez la drosophile[247]). 2/ On peut contingentement concevoir Idéalement la victoire de l'Allemagne ou des Forces Alliées à l'issue de la Seconde Guerre Mondiale. Donc, ce fait était métaphysiquement contingent[248].* | Formellement : $\neg II(p) \wedge \neg NCI(p) \rightarrow \neg MI(p) \wedge \neg MN(p)$ (en effet : pour toutes propositions q et r, telles que $q \rightarrow r$, on a : $\neg(q \rightarrow r) \leftrightarrow \neg r \rightarrow \neg q$). Or : $\neg MI(p) \wedge \neg MN(p) \leftrightarrow CCI(p)$. Donc : $\neg II(p) \wedge \neg NCI(p) \rightarrow CCI(p)$

On peut rendre encore plus intuitif ce raisonnement avec une métaphore : l'esprit est comme une sorte de laboratoire. Le philosophe maniant des expériences de pensée est

---

[246] Notons que la réciproque n'est pas nécessairement vraie. Une proposition pourrait être métaphysiquement contingente, sans être contingentement concevable Idéalement. Par exemple, bien que nous ayons considéré comme métaphysiquement impossible l'existence de cercles carrés, il se pourraient qu'ils soient métaphysiquement possibles, même si nous n'avons aucune idée de ce à quoi ils pourraient ressembler. Plus intuitivement, on peut sans doute accepter l'existence possible d'une quatrième dimension de l'espace et de créatures qui la percevraient. Négativement, on pourrait le concevoir Idéalement, mais pas Positivement. Toutefois, on peut toujours argumenter que dans ce cas, une concevabilité Idéale Positive n'est pas à notre portée d'êtres humains. Mais cela va dans le sens de ce que nous voulons montrer : le possible métaphysique peut très bien excéder le possible cognitif, à cause de nos limitations.

[247] Si on exclut le cas d'ancêtres de la drosophile qui avaient peut-être plusieurs paires d'ailes, ainsi que le cas de quelques mutations naturelles, on peut admettre qu'avant les expériences en question, il n'existait pas de mouches avec deux paires d'ailes, et qu'après ces expériences, il en a existé. Et intuitivement, il semble que ce qui n'a pas toujours existé ne saurait être nécessaire.

[248] Sous réserve de ne pas souscrire à un déterminisme métaphysique strict. Mais l'incertitude à ce sujet peut être considérée comme tranchée ou du moins tranchable par un sujet cognitif idéal.

comme un laborantin se posant des questions du genre : « Est-ce que la situation décrite par p est métaphysiquement possible ? », et se disant à lui-même : « Si ce n'est pas le cas, alors je ne devrais pas pouvoir réaliser cette situation dans mon laboratoire, car à ce moment-là, elle existerait dans au moins un endroit : mon laboratoire. Or, ce qui existe à au moins un endroit, en au moins un exemplaire, est nécessairement possible. ». Ainsi, lorsque nous nous imaginons, avec force détails, cohérence, stabilité, l'existence d'un gnome vert à deux mètres de nous, c'est un peu comme s'il existait réellement pour nous. Bien sûr, cela ne le fait pas exister réellement dans le monde physique, mais cela le fait exister réellement dans notre monde mental, lequel est soumis aux lois du monde physique, ce dernier étant soumis aux lois métaphysiques.

A ceux qui seraient sceptiques au sujet de l'expression « monde mental », nous espérons les voir moins sceptiques au sujet de l'expression « monde virtuel ». Les progrès de l'informatique rendent en effet possible une quasi-matérialisation de divers mondes concevables, avec un niveau de réalisme, de cohérence, de stabilité, de plus en plus pointu. Tant et si bien que, malgré leur vocation première dans le domaine de l'amusement, de la modélisation scientifique et technologique, de l'exploration virtuelle de reconstitutions ou de la maison de nos rêves avant d'en engager la construction, ces mondes virtuels pourraient très bien trouver leur place dans la « boîte à outils » du philosophe. L'aide à la conception d'expériences de pensée fortes et rigoureuses – basées sur une concevabilité très fortement Positive et Idéale – serait une application métaphysique potentielle très prometteuse. Loin des risques inhérents à l'à peu près plausible d'une subjectivité que d'aucuns peuvent considérer comme peu fiable, vite trompée qu'elle peut être par des rêveries sans consistance réelle, l'outil informatique donnerait au métaphysicien toute la

rigueur nécessaire à cette activité de l'esprit consistant à explorer la possibilité et la nécessité…

Les expériences de pensée ne se limitent cependant pas à tester la possibilité métaphysique de réorganisations contrefactuelles de nos concepts. Elles cherchent à évoquer, à partir de ces réorganisations, des intuitions supposées nous éclairer sur la nature profonde des choses. Dans le domaine particulier des réflexions sur l'identité personnelle, où cette stratégie est abondement employée, de vives réactions se font entendre, de la part de la communauté philosophique, sur l'efficacité d'une telle entreprise.

Nous allons voir à présent un exemple de ce scepticisme, à travers l'analyse critique de deux articles de la philosophe Tamar Szabó Gendler : "Exceptional Persons: On the Limits of Imaginary Cases"[249] (1998), et : "Personal Identity and Thought-Experiments"[250] (2002).

## III. Expériences de pensée et identité personnelle

Les expériences de pensée paraissent être de bien curieuses choses. Comment peut-on sérieusement espérer apprendre quelque chose sur le réel – à part par pure chance – à partir de l'analyse de cas fictifs ?

Pour commencer, il convient de remarquer que le fait de tirer des conséquences factuelles d'une hypothèse scientifique revient à effectuer une sorte d'expérience de pensée.

---

249 GENDLER, Tamar Szabò. "Exceptional Persons: On the Limits of Imaginary Cases"

250 GENDLER, Tamar Szabò. "Personal Identity and Thought-Experiments"

En effet, ces conséquences factuelles décrivent ce qui se passerait si telles **conditions** étaient réalisées, conditions parmi lesquelles il faut compter : la véracité de l'hypothèse en question bien entendu (que l'on cherche justement à tester), mais aussi un certain nombre de **conditions initiales** contrefactuelles, mais possibles – logiquement, métaphysiquement, physiquement, etc. Ces conséquences factuelles sont donc ce que l'on appelle des *conditionnels contrefactuels* : X n'est pas réalisé, mais si X était réalisé, alors Y serait réalisé.

Dans une hypothèse scientifique, les conditions initiales sont toutefois réalisables, au moins en principe, et les conséquences factuelles sont observables voire mesurables, au moins en principe. On a donc une stratégie qui permet directement de vérifier l'adéquation de l'hypothèse au réel.

En philosophie, les hypothèses qui donnent lieu à des expériences de pensée sont généralement **anempiriques**. Cela veut dire que leurs conséquences factuelles – quand elles en ont – sont inobservables par principe. Les expériences de pensée, à supposer que l'on puisse les réaliser concrètement – autrement dit, que l'on puisse mettre en place les conditions initiales – ne nous permettraient pas de trancher au sujet de l'hypothèse à tester si on en restait à l'empirie, car justement, les faits conséquents à la véracité de l'hypothèse ne sont pas empiriques. C'est pourquoi il est inutile de vouloir réaliser concrètement des expériences de pensée. Cela ne nous apprendrait rien de plus. Tout le problème va donc être de déterminer si l'étude *a priori* d'une expérience de pensée, l'examen de ses conséquences factuelles – mais cependant anempiriques, donc – peut nous apprendre quelque chose tout court. Comme ce n'est pas à des énoncés d'observations que l'on peut confronter une hypothèse anempirique, c'est à d'autres types d'énoncés qu'il va nous falloir la mesurer.

Il y a plusieurs manières de tester une hypothèse anempirique. On peut la confronter à des faits de nature :

**1°) Logique :** Si on se rend compte que certaines conséquences factuelles de notre hypothèse contredisent des axiomes logiques ou des conséquences logiques de ces axiomes, alors cette hypothèse est invalide. Par exemple, une hypothèse superficiellement cohérente, peut s'avérer contradictoire dans ses développements.

**2°) Théorique :** Une théorie acceptée peut en contredire une autre. Ainsi, les deux théories ne peuvent être vraies en même temps dans le même domaine de validité. C'est par exemple le cas de la mécanique quantique et de la mécanique relativiste. Pour se rendre compte de cette incompatibilité, il n'est pas besoin de réaliser des expériences, il suffit de mettre en évidence au moins un cas possible où les deux théories entrent en conflit. Il existe différents types de théories. Outre les théories scientifiques, on a aussi des théories métaphysiques comme par exemple : le temps s'écoule, le monde est objectif, la cause précède l'effet, etc. Ces théories, quoique non testables, sont cependant, parfois, des conditions de possibilité de la démarche scientifique. Or, il peut arriver que deux théories métaphysiques se contredisent (ex. : le spiritualisme et le matérialisme), ou qu'une théorie métaphysique contredisent une théorie scientifique (ex. : l'évolutionnisme et le créationnisme). Quand cette dernière est une condition de possibilité de la démarche scientifique, cela donne lieu à des débats intéressants. Ainsi, Alvin Plantinga[251] a cherché à

[251] PLANTINGA, Alvin, "An evolutionary argument against naturalism".

montrer une contradiction entre le néodarwinisme (théorie scientifique) et le naturalisme (théorie métaphysique)[252].

**3°) Conceptuel :** Une hypothèse apparemment plausible peut remettre en question nos concepts. C'est ce qui se passe avec l'identité personnelle. Les critères employés sont logiquement cohérents et scientifiquement acceptables voire plausibles. Ils ne contredisent aucun fait empirique et aucune théorie explicative sur le monde. Là où le bas blesse : ils nous conduisent à revoir notre concept de personne. Les motifs que l'on peut avancer pour privilégier un concept peuvent être de divers ordres : le concept correspond à une réalité naturelle observable (ex. : le concept de planète, de montagne, de rocher, de table, de chaise, de lac, de rivière, de caillou, d'objet, d'événements, d'espace, de temps, etc.) ; ou bien c'est un concept construit par l'homme, qui a une fonction sociale, culturelle, conventionnelle, etc. (ex. : le concept de démocratie, d'économie, de droit, d'état, de nation, de monnaie, etc.). Evidemment, il semble plus difficile de remettre en question l'existence d'une réalité ontologique correspondant aux concepts de la première sorte. C'est pourquoi il est intéressant de se demander à quelle sorte de concept, celui de personne appartient.

## *A. Le philosophe et son cantilever…*

Dans « Exceptional Persons : On the Limits of Imaginary Cases », la philosophe Tamar Gendler explique que la légitimité ou l'illégitimité de l'appel à des expériences de

[252] Seul le naturalisme méthodologique étant requis par la démarche scientifique, et le naturalisme métaphysique n'étant pas forcément adopté par les scientifiques (beaucoup se retrouvent dans un sain scepticisme eu égard à la nature ultime des choses), l'argument d'Alvin Plantinga n'est, au mieux, qu'un argument en faveur de l'agnosticisme (et plus généralement en faveur du scepticisme métaphysique de nombre de chercheurs).

pensée dans certains cas, dépend de la structure du concept que l'expérience de pensée est supposée nous aider à élucider.

- Si le concept est structuré – ontologiquement – autour d'un ensemble de **conditions nécessaires et suffisantes**, *et* si ces conditions – d'un point de vue épistémique – **contribuent concrètement à notre catégorisation ou non**, sous ce concept, d'une entité en particulier, alors les cas imaginaires peuvent nous aider à distinguer les caractéristiques essentielles vs accidentelles de ce concept.

- Si, en revanche, le concept n'est pas structuré de cette façon (et c'est le cas, selon Tamar Gendler, du concept de personne), ou si les caractéristiques en question ne gouvernent pas notre application de ce concept, alors les scénarios fictifs sont vraisemblablement trompeurs. Cela ne signifie pas pour autant que nous ne serons pas en mesure de leur donner du sens. La ressemblance, même partielle, de ces cas, avec des cas ordinaires, nous inclinera à les assimiler à ces derniers, mais la fiabilité de notre jugement ne sera alors aucunement garantie.

## 1. L'argument de Gendler

En effet, même si nous sommes conscients que deux caractéristiques ne coïncident pas nécessairement dans tous les cas possibles, le fait qu'elles coïncident dans tous les cas réels, ou même juste dans la plupart des cas réels, peut signifier tout simplement que nos concepts n'ont véritablement de sens que dans ces cas-là. Même si nous sommes capables

de *donner du sens* à des situations exceptionnelles voire fantaisistes où des caractéristiques, généralement liées, sont considérées en isolation, c'est par assimilation avec les cas ordinaires où ces caractéristiques coïncident. Mais cette assimilation, si elle peut se comprendre d'un point de vue psychologique, n'a sans doute pas de légitimité épistémique. Notre capacité à donner sens à de tels cas outrepasse notre capacité à effectuer des jugements valables à leur sujet.

Pour Tamar Gendler, cela explique à la fois l'attrait et le caractère fallacieux des expériences de pensée au sujet de l'identité personnelle. C'est un fait évolutif que les personnes humaines sont à la fois des organismes biologiques, ainsi que des consciences de soi dotées de caractéristiques psychologiques. Et tandis que ces caractéristiques sont clairement conceptuellement séparables, il n'est pas clair que nous puissions effectuer des jugements informatifs valables au sujet de ces caractéristiques considérées en isolation.

Bien que la nature nous oblige parfois à considérer certains cas bizarres dans lesquels un seul corps peut être le support de plus d'une collection d'attributs psychologiques (cas, par exemple, de personnalité multiple, d'autres sortes de dissociations ou de compartimentations, ou de pertes de mémoire puis de réapprentissages subséquents), elle ne nous présente pas de cas dans lesquels un seul ensemble de caractéristiques psychologiques peut être présent dans plus d'un corps, que ce soit diachroniquement ou synchroniquement. Et **ce fait contingent** – le fait que dans pratiquement tous les cas, un seul esprit est associé à un seul corps – **joue un rôle central dans la manière dont nous effectuons nos jugements** concernant la nature et l'importance de l'identité personnelle, puisque c'est la lunette à travers laquelle nous regardons les cas où cette coïncidence d'un seul esprit avec un seul corps ne se maintient pas.

On peut dire que ces faits contingents mais néanmoins ordinaires vont former en nous un concept dont l'usage répond bien au traitement cognitif de ces cas contingents ordinaires, mais n'est pas prévu pour d'autres usages qui s'en éloignent. Lorsqu'il y a encore de nombreux points communs avec les situations ordinaires, on peut arriver à se débrouiller encore pas trop mal et à faire un travail cognitif acceptable de ces situations. Lorsque l'on s'éloigne un peu plus de l'habituel, l'application de nos concepts ordinaires devient boiteuse. On peut, à renfort d'imagination – qui peut plus ou moins se trouver guider par un certain contexte – trouver un moyen de les faire correspondre à ces situations inédites, en les prenant par un bout ou par l'autre, mais l'entreprise s'avère périlleuse et le résultat incertain. Enfin, certaines situations sont tellement éloignées de l'habituel de nos usages de nos concepts, que ceux-ci se révèlent totalement impuissants à les appréhender.

Les concepts sont véritablement des outils, au sens propre et technique du terme, et une comparaison avec les outils concrets comme les stylos, perceuses, marteaux, etc. n'est pas inutile pour bien faire comprendre le point. Prenons par exemple un crayon. Cet outil est conçu de manière à répondre à un besoin bien précis : déposer, sur une feuille de papier, de minces lignes de graphite qui pourront représenter, ou bien des caractères, des mots, des phrases, ou bien des dessins. Il est toutefois envisageable d'obtenir un usage largement acceptable de ce genre d'outil sur d'autres matières que du papier, par exemple du bois, ou pour d'autres usages, par exemple servir de règle pour tracer des lignes droites à l'aide d'un autre crayon. On pourrait, aussi, peut-être, tenter de manger avec un crayon, s'en servir pour piquer un bout de viande ou chercher à enrouler des spaghettis autour. Si on n'a rien d'autre sous la main, où si on est dans un pays où l'usage de baguettes de bois pour manger est chose courante (par exemple en Chine), cela peut se concevoir, mais demande une certaine habileté. Si on peut, il y a sans doute d'autres outils bien mieux

adaptés à l'alimentation d'un occidental, comme une fourchette et un couteau. Enfin, il y a certaines situations pour lesquelles l'usage d'un crayon est proprement impensable : pour percer un trou dans du béton armé, pour enfoncer un clou dans une planche, pour poncer un mur, pour aller sur la Lune, etc.

## 2. Le soi et le futur

Dans la suite de son article, Tamar Gendler s'emploie à l'analyse d'une célèbre expérience de pensée que Bernard Williams présente dans « Le soi et le futur »[253]. Selon la manière dont est racontée l'expérience de pensée, ce ne sont pas les mêmes intuitions sur l'identité personnelle qui sont évoquées. Cette sensibilité de nos réponses intuitives à la forme – et non au fond – du scénario présenté, indiquerait, d'après la philosophe, que le concept de personne n'est pas de ceux qui sont – ontologiquement ou épistémiquement – structuré selon des conditions nécessaires et suffisantes et dont on peut espérer distinguer des caractéristiques essentielles vs accidentelles, par le truchement de l'analyse de cas fictifs.

Voyons cela de plus près. Dans « Le soi et le futur », Bernard Williams décrit un cas imaginaire pour lequel il est demandé de contempler une machine fonctionnant de la manière suivante : quand deux individus, A et B, entrent dans la machine, celle-ci reconfigure le cerveau de A, de manière à ce qu'il soit associé avec tous les états psychologiques précédemment associés avec le cerveau/corps de la personne B, et elle reconfigure le cerveau de B, de manière à ce qu'il soit associé avec tous les états psychologiques précédemment associés avec le cerveau/corps de la personne A. Puis, Williams demande au lecteur de considérer les deux histoires suivantes.

---

253 WILLIAMS, Bernard. "The self and the future". *In:* WILLIAMS, Bernard. *Problems of the Self.*

Dans la première, on imagine A, confronté à la perspective d'être connecté à la machine en question. De la machine émerge deux personnes : la première aura le corps précédemment associé à A, mais tous les états psychologiques précédemment associés à B ; la seconde aura le corps précédemment associé à B, mais tous les états psychologiques précédemment associés à A. Avant l'opération, on explique à A qu'une des deux personnes résultantes recevra une large récompense financière, alors que l'autre sera torturée. On demande à A de décider, sur la base d'un intérêt purement personnel, si la récompense devra être attribuée à la personne avec le corps de A ou à la personne avec le corps de B.

De l'avis de Williams et Gendler, il semble judicieux pour A d'attribuer la récompense à la personne avec le corps de B. Parmi les raisons qui semblent supporter cette décision comme étant correcte est le fait que, lorsque l'opération sera effectuée et les biens distribués, la personne avec le corps de B – mais dont les mémoires et désirs correspondent à ceux de A d'avant l'opération – dira à juste titre : « C'est tout à fait le résultat que j'ai choisi ! Et comme je suis fier d'avoir choisi ainsi ! ». Alors que, vraisemblablement, si la récompense va à la personne avec le corps de A, la personne avec le corps de B remarquera avec outrage : « Pourquoi suis-je assis ici dans un grand inconfort physique, alors que ce que je voulais était la récompense ? »[254].

L'intuition évoquée dans cette première histoire est la suivante : il semble que l'on puisse donner du sens à l'idée d'une procédure suivant laquelle deux personnes pourraient, pour ainsi dire, « échanger leur corps ».

---

254 Cf. WILLIAMS, Bernard. "The self and the future", p. 48-50.

Le second scénario est le suivant. On imagine A dans les mains d'un chirurgien fou, qui lui explique : « Demain, vous allez être le sujet de souffrances physiques atroces. Mais un peu avant que cela n'arrive, quelque chose se produira qui vous fera perdre toute mémoire et effacera tous vos souvenirs, y compris ceux de cette conversation. Puis, juste avant que d'être torturé, un ensemble différent de souvenirs vous viendrons, ceux correspondant au passé d'une autre personne. ». De cette, situation, Williams écrit : « La peur, sûrement, serait encore la réaction appropriée : non pas parce que nous ne saurions pas ce qu'il va arriver, mais parce que d'un point de vue vital au moins, nous le saurions – la torture, dont on peut en effet s'attendre qu'elle nous affectera nous-même, et qu'elle sera précédée par des dérangements mentaux. »

L'intuition véhiculée par cette seconde histoire – qui est, bien sûr, juste une variante du scénario original, mais présenté d'un point de vue unilatéral – est la suivante : il semble que l'on puisse donner du sens à l'idée d'une procédure suivant laquelle deux personnes pourraient, pour ainsi dire, « échanger leur esprit ». Notre corps semble aussi être *nous*.

Suivant la manière dont est décrite l'expérience de pensée, les intuitions qui sont suscitées sont différentes et conduisent à des conclusions incompatibles. Cet exemple semble donc bien attester la thèse de Tamar Gendler suivant laquelle notre capacité à donner du sens à des scénarios imaginaires – où des caractéristiques coïncidant dans pratiquement tous les cas réels sont recombinées d'une autre façon – outrepasse notre capacité à en faire des jugements valides. Nous pouvons bien sentir qu'un scénario est parfaitement cohérent, sans pour autant savoir ce que nous ferions ou dirions si nous le rencontrions. Dans de telles circonstances, notre évaluation dépendrait vraisemblablement de la manière dont le scénario en question nous serait présenté.

### 3. La morale des histoires

Dans une dernière partie de son article, Tamar Gendler esquisse une comparaison avec une situation bien plus ordinaire :

- Supposons que vous avez décidé d'aller au théâtre. Au moment de payer votre ticket (qui coûte 10 euros), vous vous rendez compte que vous avez perdu un billet de 10 euros qui était dans votre portefeuille. Vous avez cependant encore de l'argent sur vous pour acheter votre ticket. Décidez-vous toujours d'en acheter un ?

- Supposons, cette fois-ci, que vous aviez payé votre ticket à l'avance. Arrivé au théâtre, vous vous rendez compte que vous avez perdu votre ticket. En achetez-vous un autre ?

Des études[255] ont montré qu'une majorité de personnes répondent « oui » dans le premier cas, mais que la plupart répondent « non » dans le second cas. Pourtant, il semble clair qu'à un niveau profond, ces cas sont similaires. Ils diffèrent seulement par la forme de l'item de 10 euros qui est perdu (un billet dans le premier cas, le ticket pour le théâtre dans le second cas).

Pour Tamar Gendler, bien qu'il soit de prime abord irrationnel de traiter ces deux cas là de façon différente, chaque réponse est rationnelle en un certain sens, permissif et non

[255] TVERSKY, Amos, et KAHNEMAN, Daniel. "The framing of decisions and the rationality of choice", p. 457.

coercitif, car on peut trouver des arguments en faveur de l'une ou de l'autre, et rien ne permet de trancher objectivement de manière décisive.

Faisant sienne une analyse de Nozick[256], Tamar Gendler affirme que lors d'une prise de décision, en général, nous donnons sens à un certain scénario auquel nous sommes confronté, en l'assimilant à une classe de cas familiers. Puisqu'il est, en général, économiquement peu avisé de remplacer automatiquement ce qui est cassé ou perdu, et puisque nous traitons le cas du ticket perdu dans cette rubrique, nous sommes hésitants à en acheter un autre. Et puisque, en général, cela nous rendrait malheureux de nous interdire de l'amusement dans un domaine lorsque quelque chose va mal dans un autre, et puisque nous traitons le cas du billet de 10 euros perdu dans cette rubrique, nous sommes préparés à acheter malgré tout le ticket de théâtre. Même lorsque l'on prend conscience de la similarité de fond entre les deux cas, nous pouvons maintenir que cela fait sens de les traiter différemment, nous dit Tamar Gendler, puisque chaque cas se comprend mieux comme appartenant à une des deux classes de cas entre lesquels nous faisons une distinction justifiable.

Pour Gendler, le même phénomène se produit dans le cas de Williams :

> « [Q]uand nous essayons de donner du sens à l'histoire [de Williams], à la lumière de nos présupposés généraux au sujet des cas ordinaires d'identité personnelle, nous nous focalisons sur des problèmes impliquant le corps. Et dans les cas ordinaires, la continuité du corps assure la continuité de la personne. Aussi, lorsque l'histoire de Williams est formulée de telle manière qu'elle met au premier plan la perspective corporelle, nous considérons cette dernière comme suffisante pour la continuité de la personne. […] Formulé d'une façon qui met en relief la

256 NOZICK, Robert. *The Nature of Rationality*, chapitre 1 et 2.

> continuité psychologique impliquée, le cas est assimilé aux cas ordinaires suivant le même principe de raisonnement. Nous remarquons qu'il y a une caractéristique possédée par la personne avec le corps de B – la continuité psychologique – qui suffit pour la continuité de la personne dans les cas ordinaires. Et donc, nous sommes enclin à considérer cette caractéristique comme décisive dans ce cas également. »

Il reste une objection possible. De même que l'on pourrait peut-être défendre une théorie globale de la rationalité suivant laquelle, l'une ou l'autre des attitudes au sujet du billet/ticket de théâtre de 10 euros serait vue comme catégoriquement correcte, on pourrait peut-être défendre une théorie de l'identité personnelle suivant laquelle l'un ou l'autre des critères considérés dans l'histoire de Williams – critère corporel vs critère psychologique – capturerait réellement la vérité concernant la nature et l'importance de l'identité personnelle. Et en effet, c'est précisément le but de ce genre d'expérience de pensée : fournir des scénarios méticuleusement décrits dans lesquels les caractéristiques essentielles et accidentelles peuvent être séparées de manière à ce que nous puissions les identifier plus facilement.

Pour répondre à cette objection, Tamar Gendler propose de distinguer deux stratégies :

- La stratégie **exception-comme-scalpel** utilise les cas exceptionnels de manière à progressivement restreindre l'éventail de caractéristiques requises pour l'application d'un concept en nous permettant d'isoler les caractéristiques essentielles pour l'application du concept de celles qui sont de simples corrélats ordinaires.

- La stratégie **exception-comme-cantilever** appréhende les instances extraordinaires d'une catégorie comme essentiellement dépendantes de ses instances ordinaires, contre lesquelles, elles peuvent être vues comme des exceptions.

Maintenant, si on applique ces deux stratégies au cas de Williams, nous concluons ce qui suit :

- Suivant la stratégie exception-comme-scalpel : ni la continuité physique ni la continuité psychologique ne sont nécessaires ou suffisantes pour le maintien de l'identité personnelle à travers le temps. Pour Tamar Gendler, cela suggère que le concept de personne n'est pas organisé autour d'un ensemble de conditions nécessaires et suffisantes jouant un rôle dans l'identification des candidats tombant sous ce concept.

- Suivant la stratégie exception-comme-cantilever : la première version du cas de Williams nous montre qu'il existe des cas imaginables pour lesquels nous serions enclin à attribuer une identité diachronique en l'absence de toute continuité corporelle ; la seconde version nous montre la même chose au sujet de la continuité psychologique. Mais plutôt que d'en conclure quelque chose au sujet des conditions nécessaires et suffisantes d'application du concept de personne (ou de l'absence de ces conditions), la stratégie exception-comme-cantilever nous invite à conclure ceci : nos décisions au sujet des cas exceptionnels où une identité personnelle diachronique est obtenue sont justifiées par la permissivité rationnelle de

les assimiler à des cas ordinaires. Dans le premier scénario, on se focalise sur la psychologie. Dans le second, sur la continuité corporelle. Mais dans les deux cas, on interprète ce qui se passe en référence aux cas ordinaires.

## 4. Commentaires

Récapitulons brièvement les points du raisonnement de Gendler :

1. Même si certaines situations sont métaphysiquement, voire physiquement et conceptuellement possibles, et même si on peut donner un sens à de telles situations, il n'est pas dit que l'on puisse en faire un jugement valable. Nos concepts, dans leur genèse, leur nature, et leurs conditions d'application, sont étroitement liés à certaines caractéristiques contingentes de la réalité à laquelle nous avons à faire habituellement. En dehors de ces limites, leur usage n'est pas garanti.

2. Les expériences de pensée n'ont un pouvoir élucidant que pour les concepts structurés – ontologiquement et épistémologiquement – autour d'un ensemble de conditions nécessaires et suffisantes.

3. La sensibilité de nos réponses intuitives à la manière de présenter les expériences de pensée, telles celle de Williams, concernant l'identité personnelle diachronique, indique que le concept de personne n'est pas structuré – épistémologiquement – autour d'un ensemble de conditions

nécessaires et suffisantes. (et donc : on ne peut élucider un tel concept à l'aide d'expériences de pensée).

4. La stratégie exception-comme-scalpel appliquée aux expériences de pensée sur l'identité personnelle diachronique, telles celle de Williams, montre que le concept de personne n'est pas structuré – ontologiquement – autour d'un ensemble de conditions nécessaires et suffisantes.

5. La stratégie exception-comme-cantilever appliquée aux expériences de pensée sur l'identité personnelle diachronique, telles celle de Williams, montre que le concept de personne est plutôt, dirons-nous, de nature « prototypique »[257].

Nous nous permettrons à présent d'apporter quelques remarques et critiques à ces différents points :

### a) Nos concepts ne sont-ils fiables que dans le cadre étroit des usages qui ont contribué à leur formation ?

Dès le premier point, il semble qu'il y ait quelques problèmes. Bien sûr, nos concepts ne tombent pas d'un Ciel Platonicien d'Idées Eternelles… Ils se forment selon un processus très concret d'apprentissage. La confrontation à la réalité structure nos concepts

[257] Nous nous inspirons de la formulation que Daniel Cohnitz utilise dans sa propre analyse de l'article de Gendler. Cf. COHNITZ, Daniel. "Personal Identity and the Methodology of Imaginary Cases", p. 21-23.

d'une certaine façon qui ne les rend pas forcément toujours aptes à répondre à toutes les situations métaphysiquement possibles. Mais, il semble qu'il existe une certaine flexibilité dans cet apprentissage. Contrairement à ce que l'on a pu penser à une époque, notre esprit n'est pas une *tabula rasa* où viennent se déposer passivement les impressions sensorielles. Notre esprit a un rôle actif, et c'est précisément ce rôle actif qui permet une certaine flexibilité. A partir des données dont il dispose, il est en mesure d'effectuer un certain nombre d'opérations qui lui permette de s'adapter à la nouveauté, à l'imprévu. Ainsi, même si nous n'avons pas été confronté à toutes les instances existantes de la classe des jeux, il nous est possible, non seulement d'interpoler, mais également d'extrapoler cette classe. Par exemple, quelqu'un habitué aux jeux classiques (dames, scrabble, jeux de cartes, etc.) qui découvrirait les jeux de rôles sera peut-être quelque peu désappointé au début, certes. Mais après un certain temps à les pratiquer – et à moins d'avoir l'esprit quelque peu figé – n'aura-t-il pas incorporé cette catégorie d'activité dans la classe des jeux ? Quelqu'un n'ayant vu de sa vie que des chats gris, noir ou blanc (mais ayant vu aussi d'autres animaux pouvant avoir d'autres couleurs), et à qui l'on montrerait un animal ressemblant comme deux gouttes d'eau à ce qu'il connaît de la classe des chats, mais qui aurait un pelage roux, subsumerait très probablement ce nouveau spécimen sous le concept « chat ». Ce jugement serait-il audacieux et sans garantie ? Les concepts que nous utilisons sont, certes, adaptés aux besoins que nous avons rencontrés dans le passé, mais ils disposent d'une certaine plasticité qui nous permet de prévoir et donc d'englober un certain nombre de situations contrefactuelles qui *pou rraient* se présenter. Bien sûr, au plus ces situations contrefactuelles s'éloignent de ce à quoi nous sommes habitués, au moins nos concepts sont utilisables de manière valide, mais où situer une limite précise ? Et pourquoi ces situations contrefactuelles ne pourraient-elles pas participer précisément à l'enrichissement de nos concepts comme le font les situations factuelles ? Considérons encore ceci : le petit enfant a

déjà un certain concept d' « homme » et de « femme ». Par exemple, il pourra se dire : « Une femme est quelqu'un qui a des cheveux longs, une voix douce, une peau douce, qui porte des jupes ou des robes. Un homme est quelqu'un qui a des cheveux courts, de la barbe, une voix grave, qui porte des pantalons. ». Il va de soi qu'un tel concept ne va pas à l'essentiel et qu'il souffre de nombreuses exceptions. Plus tard, l'enfant apprendra que les femmes ont parfois les cheveux courts, mettent parfois des pantalons, qu'elles portent les enfants à naître dans leur ventre et les allaitent avec leurs seins ; que les hommes ont parfois des cheveux longs, que dans certains pays ils peuvent porter des robes, etc. Enfin, il découvrira d'autres différences encore plus essentielles, au niveau anatomique, physiologique, psychologique. Adulte, il apprendra peut-être le rôle des chromosomes sexuels dans la détermination du sexe, étudiera peut-être certains cas plus ou moins ambigus : hermaphrodites, transsexuels, travestis, homosexuels, etc. qui l'obligeront à repréciser ses concepts. De la petite enfance à l'âge adulte, on voit comment une paire de concepts – homme vs femme – s'enrichit progressivement par une confrontation progressive et naturelle avec un ensemble de situations et de besoins de plus en plus variés. Pourquoi, après avoir passé en revue, et dans le détail, un maximum de situations actuelles, devrions-nous nous dire : « C'est assez à présent. Si je me mets à réfléchir à ce cas particulier que je n'ai pas encore rencontré jusque maintenant, je vais dire des choses non pertinentes. » ? On pourrait se le dire bien plus tôt alors : « Il me suffit de savoir que les hommes ont un pénis et aiment le foot et que les femmes n'ont pas de pénis et aiment faire la cuisine. Je pourrais bien entendu imaginer un être qui aurait un pénis et qui aimerait faire la cuisine, ou un être qui n'aurait pas de pénis et qui aimerait le foot, et de fait, il m'arrive, quoique de manière exceptionnelle, d'en rencontrer. Mais ces cas rares outrepassent le domaine de validité de mes concepts que j'ai forgé sur le corpus des cas les plus habituels. Par conséquent, ce que je pourrais dire de leur identité sexuelle serait sans doute

compréhensible à certains égards, mais non pertinent d'un point de vue épistémologique. ». Une telle caractérisation suffirait sans doute aux usages d'un enfant jouant dans la cour de récréation. Elle ne peut légitimement pas satisfaire un adulte. Ce dernier pourrait se dire : « Il me suffit de savoir que les hommes ont un pénis et des chromosomes sexuels XY et que les femmes n'ont pas de pénis et ont des chromosomes sexuels XX pour l'ensemble de mes besoins actuels. Je pourrais bien entendu imaginer un être qui aurait à la fois des chromosomes XY et pas de pénis[258], mais ce que je pourrais en dire serait non pertinent. ». Mais cela ne suffirait peut-être pas aux usages d'un médecin. De même, nos concepts ordinaires de personne suffisent aux besoins de l'occidental moyen. Ils ne peuvent légitimement pas suffire aux besoins du métaphysicien en quête d'essences, de l'hindouiste qui croit à la réincarnation, du psychiatre confronté au syndrome de personnalité multiple, du neurologue qui étudie les patients split brain.

### b) Les expériences de pensée ne permettent-elles de dégager les caractéristiques essentielles vs

[258] Le syndrome d'insensibilité aux androgènes (ou syndrome du testicule féminisant) est précisément un trouble génétique qui peut avoir une telle conséquence. A la naissance, ces individus se présentent morphologiquement comme des filles et sont élevés comme tel, sans que personne ne se doute de rien. Psychologiquement et affectivement, ils/elles semblent relativement bien assumer cette éducation, présentant rarement de dysphorie du genre ou de tendances homosexuelles. Toutefois, à la puberté, leur dimorphisme sexuel s'accentue à peine et ils/elles ne sont pas réglés. Ils/elles sont également, bien sûr, stériles. C'est à ce moment là qu'un examen médical révèlera leur particularité : ils/elles sont, génétiquement, des hommes, ne présentent pas d'ovaires, mais des testicules internalisés à la place, leur « vagin », enfin, ne débouche pas sur un utérus. Se pencher sur leur identité est-il une tâche futile et sans garantie ? Pas nécessairement. En principe, suite à un tel diagnostic, il est arrivé que l'on préfère en épargner le/la patient(e) afin de ne pas le/la déstabiliser dans son identité psychologique. Ces personnes, on la dit, se comportent comme si elles étaient des femmes, passent pour des femmes dans toutes les situations de la vie ordinaire, sont répertoriées comme femmes auprès des institutions, etc. Ne sont-elles donc pas des femmes en un sens *essentiel* ?

**accidentelles des concepts que si ces derniers sont structurés autour d'un ensemble de conditions nécessaires et suffisantes ?**

Les expériences de pensée ont, à l'égard des concepts prototypiques, au moins une utilité : celle de les détecter. La stratégie de Tamar Gendler pour disqualifier l'usage des expériences de pensée au sujet de l'identité personnelle est l'analyse de deux versions différentes d'une expérience de pensée au sujet de l'identité personnelle… La philosophe ne propose pas de méthode qui se passe d'expériences de pensée pour expliciter concrètement cette distinction entre les concepts structurés autour de conditions nécessaires et suffisantes et les autres, prototypiques.

Tamar Gendler pense que les expériences de pensée peuvent nous aider à distinguer les caractéristiques essentielles vs accidentelles des concepts, à condition que ces derniers soient structurés selon un ensemble de conditions nécessaires et suffisantes et que ces conditions soient ce qui nous permet d'identifier les instances de ce concept comme telles.

Si c'est le cas, alors il semble bien que l'ensemble des concepts élucidables par des expériences de pensée est vraiment très mince. Même les concepts de la physique ne font que tendrent vers cet idéal de conditions nécessaires et suffisantes. Enfin, on peut défendre l'idée suivant laquelle les expériences de pensée, loin de ne pouvoir convenir à l'élucidation des concepts prototypiques, sont une activité parmi d'autres qui peut justement concourir à leur réorganisation autour de conditions nécessaires et suffisantes.

Précisément, la plupart de nos concepts naturels et que nous définissons et reconnaissons à partir de conditions nécessaires et suffisantes ne commencent-ils pas leur existence de manière prototypique, sans être organisés – du moins épistémologiquement – de cette manière ? C'est la confrontation à d'autres cas – actuels et, pourquoi pas, contrefactuels, mais encore faut-il déjà les accepter ! – ainsi que l'approfondissement de la connaissance des cas actuels, qui va guider l'esprit dans un travail de conceptualisation qui va le faire passer de concepts prototypiques à des concepts organisés selon des conditions nécessaires et suffisantes, qui va faire passer de la connaissance des symptômes à la détermination des critères. Prenons un exemple paradigmatique : la classification des êtres vivants.

Le contact de l'homme avec la nature l'a conduit très tôt à classer les êtres vivants en différentes catégories, selon divers besoins. Il n'est nullement nécessaire d'effectuer des études poussées pour se rendre compte que, grossièrement, il existe, par exemple, des animaux terrestres avec des poils et dont les femelles ne pondent pas d'œufs et allaitent leurs petits : les mammifères, d'autres animaux, munis d'ailes, de plumes et d'un bec, capables de voler, qui pondent des œufs et n'allaitent pas leur petits : les oiseaux, d'autres animaux encore, qui vivent dans l'eau, sont munis de nageoires et non de pattes, pondent des œufs et n'allaitent pas leurs petits : les poissons. Cette première distinction conceptuelle est bien évidemment prototypique. Une étude un peu plus poussée fait apparaître l'existence d'une série de cas problématiques : des animaux qui allaitent mais pondent des œufs (échidnés et ornithorynques), des animaux qui ont des plumes mais ne volent pas (autruches, kiwis, émeus, etc.), des animaux qui ont des poils, allaitent leurs petits, mais volent (chauves-souris), des animaux qui ne pondent pas d'œuf, allaitent leurs petits mais n'ont pas de poils, vivent dans l'eau et sont munis de nageoires (baleines, dauphins,

cachalots, etc.), des animaux qui vivent dans l'eau, sont munis de nageoires, n'allaitent pas leurs petits, mais ne pondent pas d'œufs (requins, raies, etc.), etc.

L'histoire des sciences nous apprend que les hommes se sont parfois trouvés hésitants pour classer les êtres vivants : ainsi, la chauve-souris fut parfois classée parmi les oiseaux, la baleine parmi les poissons, les coraux parmi les végétaux, etc. L'épistémologie génétique nous montre que certaines distinctions conceptuelles ne sont pas évidentes à un très jeune âge : il est difficile de faire passer l'idée selon laquelle les baleines sont des mammifères au même titre que les vaches, et pas des poissons, par exemple. Et même chez les adultes, peu sans doute sont conscients du fait que les orvets ne sont pas des serpents mais des lézards sans patte, et que les ginkgo, malgré leur feuilles, sont plus proches des conifères que des « feuillus ».

Avant la révolution apportée par la biologie moléculaire, tout le travail des systématiciens et des zoologues a été d'essayer de dégager un certain nombre de caractéristiques essentielles, afin de repréciser ces concepts. La présence de plumes s'est avérée caractéristique essentielle de la classe des oiseaux, l'allaitement des petits par les femelles s'est trouvé être une caractéristique essentielle de la classe des mammifères, la présence de branchies durant toute la vie s'est trouvée caractéristique de la classe des poissons. Cette première caractérisation s'est trouvée, ensuite, en partie confirmée par la paléontologie puis par la biologie moléculaire. La paléontologie, tout d'abord, mit en évidence l'unité phylogénétique des mammifères, descendants probables d'un groupe de reptiles du Trias : les thérapsidés ; et des oiseaux, descendants probables d'un groupe de petits dinosaures carnivores, au Jurassique. Pour les poissons, la chose est moins claire.

Leur origine se perd dans les débuts de l'ère Primaire. Leurs ancêtres directs, selon toute vraisemblance, ressemblaient à nos actuels amphioxus et ascidies.

La surprise viendra de la biologie moléculaire. Elle montra l'unité génétique des mammifères, et celle des oiseaux, confirmant les résultats de la paléontologie. Par contre, elle montrera une chose curieuse : génétiquement (et donc phylogénétiquement), les poissons ne sont pas une classe naturelle. Et bien au contraire, nous sommes plus proches de certains poissons (les sarcoptérygiens) que ces derniers ne sont proches d'autres poissons (les actinoptérygiens). Ces deux groupes sont eux-mêmes plus proches l'un de l'autre que du groupe des chondrichtyens (requins, raies). Enfin, ces trois groupes sont plus proches l'un de l'autre que du groupe des agnathes (lamproie), que les zoologues avaient déjà écartés de la classe des poissons à cause de leur absence de mâchoire articulée. De manière plus éclatante encore, on voit ainsi que le concept de poisson est prototypique. On l'applique à une diversité de cas qui partagent, certes des traits communs importants (nageoires, parfaite adaptation respiratoire à la vie marine), mais non suffisants (certains amphibiens conservent des branchies toutes leur vie, certains mammifères ont des nageoires), et finalement superficiels, au regard de l'histoire des espèces et de la biologie moléculaire. Les biologistes ont d'ailleurs abandonné le concept de poisson. A la place, à présent, on parle tout simplement des agnathes (les « poissons » sans mâchoire articulée), des chondrichtyens (les poissons cartilagineux), des actinoptérygiens (les poissons osseux à nageoires rayonnées : truite, poisson rouge, hippocampe, carpe, etc.), et des sarcoptérygiens (les poissons osseux à nageoires charnues : dipneustes et coelacanthes). Depuis peu, les amphibiens, reptiles, oiseaux et mammifères, sont regroupés aussi parmi les sarcoptérygiens, à côté des membres pisciformes de ce groupe zoologique (dipneustes et coelacanthes). A présent, et bien que la rencontre d'un cas d'école inclassable soit toujours

possible dans l'absolu, et bien que quelques querelles de spécialistes sévissent toujours, on peut dire que la classification moderne des êtres vivants a relativement bien cerné une certaine essence, et que les concepts de mammifère, d'oiseau, etc. sont désormais relativement bien définis par un ensemble de conditions nécessaires et suffisantes. Le concept de poisson, qui était irréductible à un ensemble de conditions nécessaires et suffisantes, a été tout simplement éliminé, au profit de nouveaux concepts : agnathes, chondrichtyens, actinoptérygiens, sarcoptérygiens. Ceci vaut sur un plan ontologique, puisque cette classification n'est pas qu'une pure convention linguistique reflet d'intérêts humains de base, mais semble bien correspondre à une réalité biologique et paléontologique objective au sens fort du terme. Et ceci vaut aussi sur un plan épistémologique puisque, même si le non spécialiste continue de reconnaître les animaux comme appartenant à la classe des mammifères ou des oiseaux sur la base de symptômes de surface comme la présence de plumes ou de poils, la pratique scientifique des zoologues et systématiciens se fonde sur des critères autrement plus précis et discriminants.

Que de nombreux concepts aient commencé sous une forme prototypique, avant de s'organiser selon des conditions nécessaires et suffisantes, ne prouve pas, bien entendu, qu'il en soit de même avec le concept de personne. Cela laisse juste l'option ouverte. Certains concepts pourraient peut-être être irréductiblement prototypiques. Mais comment le savoir ?

### c) L'influence du mode de présentation d'une expérience de pensée visant à élucider un concept sur nos réponses intuitives est-elle un indicateur que ce concept est organisé de manière prototypique ?

Tamar Gendler estime que les réponses intuitives aux expériences de pensée visant à élucider des concepts prototypiques sont sensibles à la présentation que l'on en donne, contrairement aux concepts organisés autour de conditions nécessaires et suffisantes.

L'exemple de la classification des êtres vivants que nous avons développé plus haut va nous permettre de fournir un contre exemple à cette thèse à travers une nouvelle expérience de pensée.

Imaginons que vous soyez un extra-terrestre appartenant à une civilisation très avancée, en visite sur Terre pour l'étudier. Nous sommes 140 millions d'années avant l'apparition de l'homme. Vous vous êtes fait spécialiste des dinosaures. Vous en avez déjà répertorié plusieurs espèces différentes que vous avez classées en différentes familles et différents genres. Un jour, vous êtes confronté à un curieux animal : un archéoptéryx. Sa principale particularité : des plumes. Bien sûr, et puisque l'on supposera que c'est la première fois que vous en voyez, vous n'appellerez pas cela des plumes. De votre point de vue à vous, qui avez beaucoup étudié les écailles des reptiles, ce ne sont jamais que des écailles modifiées. Puisque vous n'avez pas rencontré d'autres animaux de ce genre, et puisque, manifestement, le reste de son organisation anatomique rappelle beaucoup celle des dinosaures, vous en conclurez certainement que ce n'est qu'un genre un peu particulier de dinosaure.

Imaginons à présent que vous soyez un savant humain vivant dans un monde post-apocalyptique où tous les reptiles ont disparu de la surface du globe. De plus, tous les documents attestant de l'existence, autrefois sur Terre, des reptiles, ont été détruits il y a de cela plusieurs siècles. Les oiseaux, eux, ont survécu et se sont bien diversifiés. Vous vous êtes d'ailleurs fait spécialiste d'ornithologie. Un jour, vous êtes confronté à un fossile d'archéoptéryx. Certes, cet animal bizarre à des ailes griffues et des dents. Mais il a surtout des plumes. De votre point de vue d'ornithologue qui n'avez jamais entendu parler de reptiles, et encore moins de dinosaures, cela ne fera aucun doute : l'archéoptéryx est bien un oiseau.

Nous venons de montrer qu'un même animal, l'archéoptéryx, aux caractéristiques ambiguës, peut être vu, soit comme un reptile, soit comme un oiseau, selon l'expérience de pensée dans lequel on le place. Bien sûr, ce que nous avons montré est peut-être qu'après tout, le concept d'oiseau n'est pas si organisé que cela selon un ensemble de conditions nécessaires et suffisantes (mais cela reste une utilité de notre expérience de pensée que de le montrer). Ou alors, peut-être cela nous indique-t-il qu'il faudrait créer une nouvelle classe d'animaux, entre les reptiles et les oiseaux (ce qui serait aussi une conclusion utile). En fait, si, comme nous l'avons montré sur l'exemple de la classification des êtres vivants, l'évolution des concepts tend à leur faire acquérir une forme structurée autour de conditions nécessaires et suffisantes, ce processus n'est jamais totalement fini est définitif (à moins d'imposer de manière dogmatique une « vérité » absolue sur la signification des concepts).

Quel autre exemple de concept pourrions-nous trouver qui soit plus convaincant ? A vrai dire, il n'y a guère que des concepts *a priori*, comme ceux de la logique, qui peuvent se

prévaloir d'être structurés autour de conditions nécessaires et suffisantes strictes. Mais ces concepts sont déjà élucidés dès le départ, semble-t-il[259]. Dès lors que l'on s'intéresse à des concepts *a posteriori*, une telle structuration ne peut être qu'asymptotique. Et pourtant, c'est pour ces concepts que des expériences de pensée sont intéressantes, ne serait-ce que pour démontrer qu'ils ne sont pas si structurés autour de conditions nécessaires et suffisantes que l'on pourrait le penser de prime abord.

Parmi les concepts *a posteriori*, on peut penser que ce sont ceux de la physique qui sont les plus purs (comprendre par là : les mieux structurés autour de conditions nécessaires et suffisantes). Et pourtant… Considérons par exemple le concept de chaleur. La chaleur est définie comme l'énergie cinétique moyenne des molécules. Cela veut dire qu'il faut et qu'il suffit que des molécules soient en mouvement pour produire de la chaleur, et qu'inversement, à chaque fois qu'il y a de la chaleur, il y a du mouvement. Mais puisque cette identité est *a posteriori*, on peut parfaitement concevoir qu'elle soit fausse. On pourrait imaginer une situation expérimentale telle que si elle se réalisait, elle démontrerait que cette identité ne tient pas. C'est une condition *sine qua non* pour que la physique soit une science expérimentale. Si on ne pouvait pas concevoir de situations expérimentales pouvant possiblement réfuter ses hypothèses, la physique ne satisferait pas le critère poppérien de falsifiabilité. L'identité entre la chaleur et l'énergie cinétique moyenne des molécules n'est donc pas un dogme intangible de la physique. C'est un résultat très sûr mais pas absolument certain.

[259] En tout cas, on voit mal ce qu'une expérience de pensée pourrait nous apprendre à leur propos. Toute expérience de pensée doit satisfaire à un critère minimum de cohérence logique. Et, pour des raisons de circularité, l'on ne peut pas élucider un concept avec l'aide de concepts dont le concept que l'on veut élucider fait partie. Si, dans le cas de Williams, on postule dès le départ que l'identité personnelle repose sur la continuité psychologique, alors on ne pourra pas démontrer autre chose, dans chacune des versions de l'histoire, que c'est la continuité psychologique qui a assuré le maintien de l'identité personnelle.

Une expérience de pensée relative à un concept physique pourrait-elle être sensible à la manière dont nous la présentons ? Nous allons essayer de répondre à la question suivante par une expérience de pensée : de l'étendue et de la masse, y en a-t-il une des deux qui est une caractéristique essentielle de la matière tandis que l'autre serait accidentelle ?

Expérience de pensée : On supposera que l'on dispose du pouvoir de comprimer ou d'étirer à volonté les objets.

La masse est essentielle, l'étendue est accidentelle : Prenons un objet de masse *m* et de volume *v*. Etirons le mentalement. Il va gagner en étendue, mais sa masse sera constante. A un point limite, son étendue sera infinie mais sa masse sera toujours la même. Comprimons le mentalement. Il va perdre en étendue mais sa masse restera constante. A un point limite, son étendue sera nulle mais sa masse sera toujours la même.

L'étendue est essentielle, la masse est accidentelle : Prenons un objet de masse *m* et de volume *v*. Etirons-le mentalement. Il va gagner en étendue et sa masse restera constante. Mais à un volume de 2 *v*, coupez le en deux et refaites deux objets de volume *v*. Chacun aura une masse de ½ *m*. Répétez le processus de nombreuses fois pour chaque objet. On obtient ainsi *n* objets de volume *v* et de masse (1/*n*) *m*. Lorsque *n* tend vers l'infini, la masse de chaque objet, (1/*n*) *m*, tend vers 0, alors que le volume de chaque objet, *v*, reste constant. Prenons *n* objets de masse *m* et de volume *v*. Comprimons les mentalement. Ils vont perdre en étendue mais leur masse restera constante. A un volume de ½ *v*, prenez deux objets pour en refaire un seul de volume *v*. Chacun aura une masse de 2 *m*. Répétez le processus *n* fois. On obtient ainsi 1 seul objet de volume *v* et de masse *n.m*. Lorsque *n* tend vers l'infini, la masse de l'objet, *n.m*, tend vers l'infini, alors que son volume, *v*, reste constant.

La première version de l'expérience de pensée suggère qu'un objet d'étendue nulle, finie ou infinie peut avoir une masse finie et constante, et donc que c'est la masse qui est la propriété essentielle de la matière, l'étendue n'étant qu'accidentelle. La seconde version de l'expérience de pensée suggère totalement l'inverse, à savoir qu'un objet de masse nulle, finie ou infinie peut avoir une étendue finie et constante, et donc que c'est l'étendue qui est la propriété essentielle de la matière, la masse n'étant qu'accidentelle.

Si l'on devait suivre les conclusions de Tamar Gendler, il semble que l'on devrait donc en conclure que le concept de matière est prototypique, qu'il n'est donc pas organisé autour de conditions nécessaires et suffisantes, et par conséquent que l'on ne peut utiliser d'expériences de pensée en vue d'en dégager les propriétés essentielles vs accidentelles.

### d) Les stratégies exception-comme-scalpel et exception-comme-cantilever montrent-t-elles que le concept de personne n'est pas structuré – ontologiquement – autour d'un ensemble de conditions nécessaires et suffisantes, que c'est un concept prototypique ?

Si oui, alors elles montrent aussi, en l'appliquant à l'expérience de pensée précédente, que le concept de matière n'est pas structuré – ontologiquement – autour d'un ensemble de conditions nécessaires et suffisantes, que c'est un concept prototypique. L'implausibilité de cette conséquence pourrait suffire à disqualifier cette stratégie. Néanmoins, nous croyons possible de démontrer qu'aucun concept, fondamentalement, n'est prototypique.

Tamar Gendler nous dit :

> « [S]upposons que les entités qui tombent sous un certain concept ont généralement les caractéristiques *a*, *b*, *c*, *d* et *e* […] et supposons de plus qu'une certaine entité soit trouvée qui a seulement *b* et *d*, mais qui tombe néanmoins sous le concept. Selon [la stratégie exception-comme-cantilever], la chose correcte à dire au sujet de l'entité en question est qu'elle tombe sous le concept seulement parce qu'elle est similaire, de certaines manières cruciales, à des instances plus typiques d'entités que la théorie décrit. »

En quoi consiste cette similarité ? Dans l'exemple donnée, il semble que ce soit : « l'entité en question possède au moins deux caractéristiques parmi *a*, *b*, *c*, *d*, et *e*. »

Il semble donc que l'on peut reformuler la définition du concept en question comme suit : une entité est subsumée par ce concept, *si et seulement si* elle possède **au moins deux caractéristiques parmi les suivantes : *a*, *b*, *c*, *d*, et *e***. C'est bien de l'établissement de conditions nécessaires et suffisantes, dont il s'agit.

Dans le cas de Williams, on pourrait, de même, dire que : L'identité personnelle entre une personne A vivant en t1 et une personne B vivant en t2 se maintient, *si et seulement si*, il y a au moins un de ces critères qui est vérifié, de A vers B : continuité corporelle, continuité psychologique ; et *si et seulement si*, il n'y a pas d'autre personne C, coexistante de B, telle que de A vers C, au moins un de ces critères est également vérifié.

Evidemment, cela pose le fameux problème de Parfit : « Pourquoi un double succès serait-il un échec ? », qui nous conduit à penser que cette définition de l'identité personnelle

n'est pas parfaite – à l'instar de nombreux autres concepts, dont certains paraissent pourtant bien élucidés : le concept d'oiseau, le concept de matière, etc.

Mais il existerait peut-être une méthode – quasiment un « algorithme » – qui pourrait permettre, pour tout concept, de le faire tendre vers une définition sous la forme d'un ensemble de conditions nécessaires et suffisantes. Effectuée par un sujet cognitif idéal, cette méthode aboutirait à une définition stricte de chaque concept sous cette forme.

Cette méthode consisterait à répertorier, dans un maximum de situations actuelles, l'usage que les personnes font des concepts, ainsi que, pour les situations contrefactuelles, l'usage qu'elles disent qu'elles en feraient. Dans chaque cas, il faudrait aussi mesurer le plus finement l'état de leur cerveau, de leur corps, de l'environnement. Il est probable, si le monde est déterministe – rigoureusement ou statistiquement – que l'on finira par établir ainsi des conditions factuelles précises (CF) qui, si elles sont réalisées, entraînent automatiquement telle ou telle réponse (R) de la part des personnes, lorsque telle situation (S) – actuelle ou contrefactuelle – se présente ou leur est racontée. On pourra ainsi établir, pour chaque concept étudié, une équation compliquée de forme générale $R = f(CF, S)$. Cette équation donnera une définition du concept en question en termes de conditions nécessaires et suffisantes.

Appliquée au concept de personne, elle permettrait, peut-être de dire que, en moyenne, et avec un écart-type faible, une personne P dira d'une personne B vivant en t2 qu'elle est la même personne qu'une personne A vivant en t1, si et seulement si :

- B est dans une relation de continuité corporelle ou psychologique avec A, et est le seul dans ce cas là.

- Ou : B est dans une relation de continuité corporelle ou psychologique avec A, n'est pas le seul dans ce cas là, mais P est dans l'état E1.

- Ou : B est dans une relation de continuité corporelle ou psychologique avec A, n'est pas le seul dans ce cas là, mais l'environnement de B est dans l'état E2.

- Etc.

Autrement dit, ce que cet argument veut dire, c'est que la connaissance de tous les faits objectifs de l'Univers pourrait entraîner la connaissance de la vérité métaphysique concernant l'identité personnelle de chaque individu.

### e) En conclusion

Nous avons montré que nos concepts disposent d'une grande flexibilité, qui leur permet de nous garantir des jugements valables dans des situations qui outrepassent ce dont nous avons fait l'expérience. Nous avons montré également que le nombre de concepts structurés autour d'un ensemble de conditions nécessaires et suffisantes est bien plus restreint que ce que l'on serait tenté de croire. Cela a pour conséquence que si les expériences de pensée ne sont efficaces que pour dégager l'essentiel de l'accidentel pour les concepts structurés de la sorte, alors elles sont inapplicables dans un grand nombre de

domaines où on les accepte pourtant relativement mieux en règle générale. On remarque d'ailleurs un effet possible de la présentation des expériences de pensée sur notre manière de les évaluer, avec des concepts pourtant considérés comme bien structurés autour d'un ensemble de conditions nécessaires et suffisantes. Enfin, nous avons essayé d'étayer l'idée suivant laquelle, fondamentalement, aucun concept n'est prototypique.

Cette critique générale de la méthodologie des expériences de pensée, lorsqu'elles sont appliquées au concept de personne, concernait surtout la validité des intuitions véhiculées par les scénarios imaginaires. Dans l'article que nous examinerons à présent, Gendler, reconnaissant que certaines expériences de pensée sur l'identité personnelle conduisent à des intuitions valables, s'attaque cette fois-ci au problème de leur généralisation.

### *B. Des désaccords sur la méthode de l'accord*

Nos *jugements* concernant des ***scénarios imaginaires*** à travers lesquels on cherche à analyser un **concept**, sont vraisemblablement *peu fiables* quand (1) les scénarios impliquent des *perturbations* de certaines **corrélations contingentes** qui jouent un rôle central dans notre *compréhension* et *utilisation* du **concept** étudié, et quand (2) le **concept** en question concerne l'*évaluation d'une valeur* (et non d'un fait). Les caractéristiques qui *expliquent* ou *justifie* une relation **peuvent être distinctes** des caractéristiques qui l'étayent en tant que *conditions nécessaires*. Voilà la thèse que cherche à défendre Gendler dans un autre article, paru en

2002 : "Personal Identity and Thought-Experiments"[260]. Elle se base, pour cela, sur l'analyse de l'argument de la fission, selon Parfit, et de l'expérience de pensée qui l'étaye[261].

La philosophe commence par dégager ce qu'elle croit être trois présupposés cruciaux de l'argument de la fission :

a) Principe d'intrinsécisme : La relation d'importance (M) pour le souci prudent rationnel (RPC) est une relation intrinsèque.

b) Principe de suffisance : Si la relation M se maintient entre A et B, alors le souci prudent de A pour B est rationnel : $M(A,B) \rightarrow RPC(A,B)$

c) Principe de nécessité : Si le souci prudent de A pour B est rationnel, alors la relation M se maintient entre A et B : $RPC(A,B) \rightarrow M(A,B)$

Elle explique ensuite qu'il existe deux sens de M qui peuvent nous intéresser :

(1) Le sens « facteur commun » : Rechercher quelle relation (le cas échéant) est commune à tous les cas (possibles) où RPC est obtenu. Et selon ce sens-là, a), b), et c), sont VRAIS.

(2) Le sens « explicatif » : Rechercher la relation qui explique pourquoi RPC est obtenu. Et dans ce cas-là, ni a) ni c) ne sont VRAIS.

---

260 GENDLER, Tamar Szabò. "Personal Identity and Thought-Experiments"

261 PARFIT, Derek. *Reasons and Persons*, p. 254-255.

## 1. L'argument de la fission selon Parfit :

Gendler présente ensuite l'expérience de pensée de Parfit comme suit :

> « Des triplets sont impliqués dans un accident, dans lequel le corps d'un, Brainy, est mortellement blessé, tandis que les cerveaux de ses deux frères sont totalement détruits. Brainy est tel que les bases physiques pour ses caractéristiques psychologiques sont réalisées en double, un ensemble complet dans chaque [hémisphère[262]]. A la suite de l'accident, les docteurs divisent son cerveau en deux, et transplantent les deux hémisphères dans les corps des deux frères. »

A partir de là, nous avons deux scénarios :

1) Le cas du transfert unique, que Gendler décrit en ces mots :

> « [S]eule la greffe gauche prend, et la greffe droite est détruite. L'individu résultant, Lefty (ce terme est une abréviation pour « l'individu qui a l'hémisphère gauche[263] original de Brainy »), a toutes les mémoires et tous les états psychologiques de Brainy et un corps pratiquement indiscernable de celui que Brainy avait avant l'accident. »

2) Le cas du double transfert, qu'elle présente comme suit :

> « Ici les deux greffes prennent. Chacun des deux individus résultants, Lefty et Righty, a toutes les mémoires et tous les états psychologiques de Brainy et un corps pratiquement indiscernable de celui que Brainy avait avant l'accident. »

---

[262] Tamar Gendler note « lobe » (le mot anglais est le même que le mot français), ce qui nous apparaît peu approprié d'un point de vue anatomique…

[263] « Left » en anglais.

Le raisonnement de Parfit, pour Gendler, est le suivant :

1. Dans le cas du transfert unique, Lefty *est* Brainy. (prémisse)

2. Si A est identique à B, alors la relation qui compte pour le souci prudent rationnel, M, se maintient entre A et B. (prémisse)

3. Dans le cas du transfert unique, M se maintient entre Brainy et Lefty. (de 1 et 2)

4. La relation de Brainy à Lefty est intrinsèquement la même entre le cas de transfert unique et celui de double transfert. (prémisse)

5. *Principe d'intrinsécisme (pour M) :* M est une relation intrinsèque. (prémisse)

6. *Le résultat de parité (pour M) :* si M se maintient entre Brainy et Lefty dans le cas du transfert unique, alors M se maintient entre Brainy et Lefty dans le cas du double transfert. (de 4 et 5)

7. M se maintient entre Brainy et Lefty dans le cas du double transfert. (de 3 et 6)

8. Dans le cas du double transfert, Lefty *n'est pas* Brainy. (prémisse)

9. Dans le cas du double transfert, M n'est pas la relation d'identité. (de 7 et 8)

10. Dans le cas du transfert unique, M n'est pas l'identité. (de 6 et 9)

11. *La conclusion de l'absence d'importance de l'identité :* M n'est pas l'identité. (généralisation de 10)

Gendler s'emploie ensuite à effectuer trois distinctions cruciales. Entre, d'une part :

(a) M(A,B) : La relation qui compte pour le souci prudent rationnel et qui se maintient entre A et B.

(b) RPC(A,B) : La rationalité du souci prudent de A pour B.

D'autre part, entre :

(c) Principe de nécessité : Si le souci prudent de A pour B est rationnel, alors la relation qui compte pour le souci prudent rationnel existe entre A et B. C'est-à-dire : RPC(A,B) → M(A,B)

(d) Principe de suffisance : Si la relation qui compte pour le souci prudent rationnel existe entre A et B, alors le souci prudent de A pour B est rationnel. C'est-à-dire : M(A,B) → RPC(A,B)

Et pour finir, entre :

(e) Lecture selon le facteur commun (pour M) : *la relation qui compte pour la rationalité du souci prudent* est une relation qui est commune à tous les cas où le souci prudent rationnel est obtenu, et seulement à ces cas.

(f) Lecture explicatrice (de M) : *la relation qui compte pour la rationalité du souci prudent* est une relation qui explique la rationalité du souci prudent obtenu dans les cas où un souci prudent rationnel est obtenu.

Gendler affirme ensuite que :

> « L'argument de Parfit fonctionne si M est compris suivant la lecture « facteur commun » ; avec cette lecture, le principe d'intrinsécisme est vrai, le résultat de parité en découle, et la conclusion de l'absence d'importance de l'identité est de ce fait établie. Mais pour les objectifs de Parfit, la lecture « facteur commun » est inadéquate. Selon cette lecture, la conclusion de l'absence d'importance de l'identité dit simplement que l'identité n'est pas commune à tous les cas où il y a un souci prudent rationnel. Mais Parfit a besoin de plus que cela : si nous répondons au cas dans le sens qu'il espère, il pense que nous devrions changer nos vues concernant ce qui justifie notre souci prudent à l'égard de nos « soi » futurs. Pour montrer ceci, il a besoin de montrer que l'identité n'explique pas la rationalité de notre souci prudent. Mais lorsque M est compris dans ce sens […], le principe d'intrinsécisme est faux […]. Et si c'est ainsi, alors l'argument de Parfit ne réussit pas à démontrer ce qu'il devait démontrer. »

## 2. Pourquoi l'argument de la fission paraît si convaincant ?

D'après Gendler, si le cas de la fission paraît si convaincant de prime abord, c'est qu'il repose sur une méthode de raisonnement scientifique correct, mais que Parfit emploierait de manière inappropriée. Cette méthode, qui a été appelée « méthode de l'accord » par John Stuart Mill, est définie par le fondateur de l'utilitarisme de la façon suivante :

> « Si deux ou plusieurs instances du phénomène sous investigation ont seulement un détail en commun, le détail dans lequel seul toutes les instances s'accordent, est la cause (ou l'effet) du phénomène donné »[264]

Gendler nous propose un exemple, pour mieux faire comprendre la méthode : le cas allumette-et-incantation :

> « Supposons que à chaque fois que je frotte une allumette contre le côté d'une boîte d'allumette et que je dis : « Que la lumière soit », l'allumette s'allume ; à chaque fois que je frotte l'allumette contre le côté d'une boîte d'allumette et que je ne dis rien, l'allumette s'allume ; chaque fois que je tiens seulement l'allumette en l'air et que je dis « Que la lumière soit », l'allumette reste éteinte ; et chaque fois que je ne frotte pas l'allumette et ne récite pas l'incantation, l'allumette reste éteinte. On a :
>
> *Cas allumette-et-incantation*
>
> A = frotte l'allumette contre la boîte
> B = profère « Que la lumière soit »
> P = l'allumette s'allume
>
| | A | non-A |
|---|---|---|
| B | P | non-P |
| non-B | P | non-P |
>
> Un rapide coup d'œil au tableau montre que P est obtenu si et seulement si A est obtenu, et que l'obtention de P est indifférent à l'obtention de B. En supposant certaines conditions – que A et B sont les seuls facteurs pertinents dans ce cas ; que A entraîne P de la même manière dans les cas A-seul et A-plus-B ; que A, B et P sont spécifiés à des niveaux de descriptions similaires, etc. – cette configuration nous donne de bonnes raisons de conclure que A explique l'obtention de P, et que B ne l'explique pas. »

Si on essaye d'appliquer la méthode de l'accord à l'argument de la fission, cela donne ceci, selon Gendler :

A = la relation R est obtenue entre X et Y
B = la relation R n'est pas obtenue entre X et quelque non-Y que se soit qui co-existe avec Y
P = le souci prudent de X pour Y est rationnel

| | A | non-A |
|---|---|---|
| B | P | non-P |
| non-B | P | non-P |

[264] MILL, John Stuart. *L'Utilitarisme*.

La philosophe inventorie ensuite un certain nombre de disanalogies qu'elle perçoit entre la méthode de l'accord, appliqué au cas allumette-et-incantation, et cette même méthode, appliquée au cas de la fission. Ces disanalogies sont les suivantes :

(a) Sujet de préoccupation : On a affaire à une explication *causale*, dans le cas allumette-et-incantation (on cherche le mécanisme effectif qui produit la lumière), alors que dans le cas de la fission il s'agit d'une explication de *valeurs* (on cherche à savoir s'il est rationnel ou non, pour une certaine personne, de se soucier prudemment d'une autre, supposée lui être ou non identique, mais en tout cas reliée à elle par une relation R).

(b) Structure interne : Dans le cas allumette-et-incantation, on s'intéresse à la contribution de deux caractéristiques indépendantes, alors que dans le cas de la fission, le cas A-plus-B est un cas particulier du cas A (être l'unique continuateur selon une relation R est une manière d'être un continuateur selon une relation R).

(c) Les conditions d'arrière-plan : Dans le cas allumette-et-incantation, les facteurs étudiés se présentent souvent de manière indépendante, alors que dans le cas de la fission, nous nous intéressons à une situation où les facteurs étudiés coïncident dans tous les cas actuels.

Puisque, dans le cas de la fission, le cas A-plus-B est un cas particulier du cas A, nous présenterons, pour plus de commodité, le cas de la fission comme suit :

| A = la relation R est obtenue entre X et Y | A | non-A |
|---|---|---|

B = la relation R n'est pas obtenue entre X et quelque non-Y que se soit qui co-existe avec Y
P = le souci prudent de X pour Y est rationnel

| | | |
|---|---|---|
| B | P | |
| non-B | P | non-P |

Pour Tamar Gendler, ce qui rend la méthode de l'accord inapplicable dans le cas de la fission – alors qu'elle est applicable dans le cas allumette-et-incantation – c'est la conjonction de (a) et (b).

S'ensuit une longue démonstration où Gendler montre que dans les cas semblables à celui de la fission pour lesquels la seule dissemblance d'avec le cas allumette-et-incantation réside en (a) ou bien en (b), la méthode est concluante. C'est vraiment, pour la philosophe, la conjonction de (a) **et** de (b) qui rend la méthode inefficace.

Exemple : le cas de Jean-Pierre, militant contre l'élevage industriel.

A = X ressemble à un morceau de viande issu d'une vache sujette à des conditions d'élevage industriel
B = X est un morceau de viande issu d'une vache sujette à des conditions d'élevage industriel
P = X est quelque chose que je ne dois pas manger

| | A | non-A |
|---|---|---|
| B | P | |
| non-B | P | non-P |

Ce qui rend ce genre de cas non concluant, pour Tamar Gendler, c'est que lorsque l'on a affaire à l'explication de valeurs, il est possible « d'élargir » à partir des cas centraux ou idéaux (par ex. : viande d'élevage industriel) vers les cas périphériques (par ex. : viande qui ressemble à de la viande d'élevage industriel sans en être), d'une manière à laquelle la méthode de l'accord est indifférente. En clair, la méthode ne nous permet pas de savoir si c'est le cas spécial qui explique le cas général ou si c'est le cas général qui explique le cas spécial.

Pour le savoir, Tamar Gendler suggère un test, qu'elle qualifie « d'association-dépendance » : si nous n'avions aucun sens qu'il puisse exister des cas comme le cas spécial (par ex. : si Jean-Pierre était ignorant des conditions de vie dégradantes des vaches dans l'élevage industriel), obtiendrions-nous encore P dans le cas général ? (par ex. : Jean-Pierre serait-il un militant contre l'élevage industriel, qui refuse de consommer toute viande qui semble en provenir ?). Ce test suggère que c'est le cas spécial et non le cas général qui explique notre évaluation.

Que se passe-t-il si nous appliquons le test d'association-dépendance au cas de la fission ? Voyons ce qu'en dit Tamar Gendler :

> « [S]upposons que nous n'ayons aucune idée qu'il puisse y avoir de l'identité, opposé à la simple relation de continuité R – y aurait-il encore une chose telle que le souci prudent ? Je suggère que la réponse est « Non ». Le concept de souci prudent est inséparable des concepts d'honnêteté, de responsabilité, de justice et de rationalité. Nos vues concernant les sortes d'obligations morales et rationnelles que nous avons envers nous-mêmes et les autres, considérés comme des êtres qui existent à travers le temps, reposent sur l'hypothèse que chacun d'entre nous avons au plus un continuateur, et que ce continuateur est quelqu'un avec qui nous serons identiques […]. Une perturbation de cette hypothèse d'arrière-plan résulterait en une perturbation de la structure entière par laquelle nous donnons sens à ce large panel de concepts. […] Et dans la mesure ou le souci prudent est interconnecté avec eux, il serait perturbé aussi. »

Toutefois, Tamar Gendler admet que dans le cas proposé par Parfit, il est rationnel pour Brainy de ressentir un souci prudent envers Lefty et Righty. Mais l'explication de Parfit, dit-elle, est erronée. Ce n'est pas parce que ce serait la relation R qui compterait et non pas l'identité qu'il est rationnel, pour Brainy, de se soucier prudemment de Lefty et

Righty, c'est parce que être lié à quelqu'un par une relation R est très ressemblant au fait de lui être identique.

Il s'agit, selon les mots de Gendler, d'une forme d'*adaptation locale*. Nous maintenons nos hypothèses d'arrière-plan concernant la continuation en général, et adaptons nos réponses lorsqu'un cas exceptionnel se présente.

### 3. Objections à Gendler

#### a) Quand Parfit affirme : « L'identité n'est pas ce qui compte », veut-il dire qu'elle ne compte pas en pratique ou qu'elle ne compte pas en théorie ou les deux ?

Pour comprendre et prédire le mouvement des objets ordinaires, on utilise, en pratique, la mécanique newtonienne. Pour comprendre et prédire le mouvement d'objets ultramicroscopiques, on utilise, en pratique, la mécanique quantique. Toutefois, alors que l'on peut prédire les résultats de la mécanique newtonienne à partir de la mécanique quantique, on ne peut pas prédire les résultats de la mécanique quantique à partir de la mécanique newtonienne. Cette dernière a un domaine de validité plus limité. D'un point de vue métaphysique réaliste, cela veut tout simplement dire que la mécanique quantique est plus « vraie » que la mécanique newtonienne, elle modélise mieux la réalité telle qu'elle est. Ainsi, l'on voit par cet exemple que, d'une part, du simple fait qu'un modèle de l'univers compte en pratique, il ne s'ensuit pas qu'il compte en théorie, ou, si l'on préfère, que son utilité dans certains cas n'implique pas sa vérité métaphysique. Sans doute, la théorie

newtonienne est vraie ou relativement vraie en ce qui concerne les objets macroscopiques, mais seulement dans la mesure où elle peut être comprise comme un cas particulier – quoique le plus fréquent, en pratique – de la mécanique quantique.

Si la relation R de Parfit n'avait rien à voir avec l'identité, Tamar Gendler aurait raison de contester la généralisation de son expérience de pensée de la fission aux cas ordinaires. Seulement voilà, la relation d'identité n'est qu'un cas particulier – quoique le plus fréquent, en pratique – de la relation R. D'un point de vue métaphysique réaliste, la relation R est donc plus profonde, plus vraie que la relation d'identité – c'est du moins ainsi que l'on peut comprendre la conclusion de Parfit. Elle a un primat ontologique sur la relation d'identité. Cette dernière peut se dériver de la relation R, en adjoignant un certain nombre de conditions supplémentaires – celles que l'on rencontre dans les cas ordinaires justement – tandis que la relation R, au contraire, ne se déduit pas de la relation d'identité. Concrètement, A vivant en t1 sera dit identique à B vivant en t2, si B est relié par une relation R à A, et s'il n'existe pas d'autres individus, contemporains de B, qui sont reliés par une relation R à A.

Ce que l'expérience de pensée de Parfit ne réfute pas (mais ne prétend peut-être pas réfuter), c'est que la relation d'identité compte, en pratique. Car en effet, dans tous les cas actuels, chaque fois qu'il y a une relation d'identité, un souci prudent est rationnellement justifié. Ce que montre (ou cherche à montrer) l'expérience de pensée, c'est que si la relation d'identité compte en pratique, ce n'est pas (ou pas tant) parce qu'elle est une relation d'identité, que parce qu'elle est avant tout une relation R, comme le montre le fait que – avec ou sans identité – du moment qu'on a une relation R, un souci prudent est toujours justifié rationnellement. Peut-être que pour mettre bien en évidence ce point, un

troisième scénario serait utile, dans lequel on n'aurait ni identité, ni relation R sans identité, et pour lequel on montrerait qu'un souci prudent ne se justifie pas.

Ce que cherche à faire Parfit, c'est à trouver un principe plus large que l'identité, qui englobe cette dernière, et qui justifie le souci prudent à un niveau plus fondamental (pour davantage de cas possibles). A travers la relation R, Parfit pense trouver ce principe plus large, englobant et fondamental. L'intérêt n'est pas tant de remplacer, en pratique, notre notion d'identité et ce qu'on y rattache, pour la remplacer par la relation R, mais de montrer que, sur un plan ontologique, il existe un primat de la relation R sur la relation d'identité. Puisque le domaine de validité de la relation R est plus large que le domaine de validité de la relation d'identité, puisque la relation R explique ou justifie tout ce que la relation d'identité explique ou justifie, et puisque, enfin, la relation d'identité peut être réduite à un cas particulier de la relation R sous certaines conditions – il n'y a donc pas de véritable compétition entre deux types de relations qui seraient irréductibles l'une à l'autre, ni d'élimination de l'une par l'autre – alors cette conclusion fait sens.

C'est donc bien l'identité qui compte en pratique, mais elle ne compte pas en théorie – ou sur un plan ontologique – car du point de vue le plus général de la réalité métaphysique, il n'y a pas, à proprement parler, d'identité, mais seulement des relations R – qui a un niveau moins profond se déploient en différentes sortes, dont des relations d'identité.

### b) La lecture en facteur commun est-elle inadéquate dans le cas de Parfit ?

Ces considérations font que, contrairement à ce qu'affirme Gendler, la lecture en facteur commun n'est pas forcément inadéquate pour les besoins de Parfit. C'est en cherchant ce qui est commun à tous les cas où un souci prudent rationnel est obtenu, et seulement à ces cas, que l'on peut dégager, par-delà la relation d'identité, le noyau dur de ce qui est essentiel à la rationalité du souci prudent. Du moins si on comprend cette rationalité en un sens épistémique.

Tamar Gendler pense que Parfit doit « montrer que l'identité n'explique pas la rationalité du souci prudent ». Il faut distinguer rationalité pratique et rationalité épistémique. Par rationalité pratique, il faut comprendre les processus, en partie cognitifs, qui nous permettent d'agir efficacement dans le monde réel. Du point de vue d'une rationalité pratique, c'est bien l'identité qui explique la rationalité du souci prudent. Mais on peut penser que c'est parce que l'identité joue le rôle d'un indicateur d'une relation M sous-jacente qui serait donc, pour Parfit, la relation R.

De même, du point de vue d'une rationalité pratique, un ciel très gris est indicateur de pluie imminente, et il est donc rationnel, lorsque le ciel vire au gris et que l'on s'apprête à sortir, de prendre son parapluie avec soi.

Par rationalité épistémique nous entendrons la compréhension profonde des mécanismes effectifs qui se mettent en place dans telle ou telle situation. Par exemple, quelqu'un pourrait savoir qu'un ciel gris est signe de pluie mais penser que la pluie vient de

fissures dans la voûte céleste qui laisseraient s'écouler une partie du contenu d'un océan gigantesque entourant la Terre. Dans ce cas là, il y aurait dissociation entre une rationalité pratique efficace et une rationalité épistémique plutôt dévoyée ou du moins non suffisamment exercée…

Il n'est pas dit que la rationalité pratique se réduise à la rationalité épistémique. Un météorologue qui aurait remarqué que de temps en temps, un ciel gris ne coïncide pas avec de la pluie, pourrait choisir de ne prendre son parapluie qu'à l'issu d'un long processus de mesures hygrométriques, pluviométriques, thermiques, bathymétriques, etc., de calculs compliqués sur ces mesures, de simulations informatiques, de vérifications, etc. Cela lui donnerait sans doute une justification supérieure au fait de prendre ou non son parapluie pour sortir et cela diminuerait sans doute son risque d'erreur, mais en pratique, se serait fastidieux, coûteux en temps, argent, énergie, puissance, etc.

La rationalité du souci prudent est, semble-t-il, le plus souvent une rationalité pratique, dans la mesure où, concrètement, lorsque l'on décide de se préoccuper de son avenir, ce n'est pas – généralement – suite à une longue théorisation philosophique sur l'identité personnelle. Nous l'avons dit, Parfit ne démontre pas que, en pratique, c'est autre chose que l'identité qui explique la rationalité de notre souci prudent. Mais qu'en est-il, du point de vue de la rationalité épistémique ?

Les rationalités épistémique vs pratique du souci prudent demandent des justifications différentes. La rationalité épistémique exige, de la part de l'attitude ou de la proposition à justifier, une base qui soit objective, c'est-à-dire indépendante du sujet et publiquement descriptible. Autrement dit, ce qui doit fonder, *de jure*, la rationalité du choix

prudent doit être, en un certain sens, extérieur au sujet cognitif. Ce doit être quelque chose qui existe dans le monde, et est valable pour tous les sujets de la même façon. C'est également quelque chose qui doit pouvoir être appréhendable par tout le monde de la même manière, pas quelque chose qui n'existe que pour certains mais pas pour d'autres. Or, si l'on suit ces desiderata, on en vient à identifier cette base qui fonde la rationalité épistémique du souci prudent – autrement dit la relation M – à quelque chose qui est commun à tous les cas où le souci prudent rationnel est obtenu.

La lecture en facteur commun n'est donc pas inadéquate, à partir du moment où l'on cherche à justifier le souci prudent rationnel selon une rationalité épistémique.

### c) Qu'est-ce qui rend la relation R et la relation d'identité si ressemblantes ?

Tamar Gendler soutient que si le souci prudent de Brainy pour Lefty et Righty est rationnel, ce n'est pas parce que ce serait la relation R qui compterait et non pas l'identité, c'est parce que être lié à quelqu'un par une relation R est très ressemblant au fait de lui être identique. Mais en vertu de quoi ces deux types de relations sont ressemblants sinon parce qu'il ont une structure commune ? Et quelle est cette structure commune, sinon une relation R ? En effet, nous avons vu qu'une relation d'identité pouvait se définir comme une relation R à continuateur unique. Et ce qui est commun à une relation R et à une relation R à continuateur unique est évidemment une relation R. n'est autre que la relation R elle-même. C'est donc bien la relation R qui compte, ultimement, même si, en pratique, on l'approxime par une relation d'identité.

En arguant de la ressemblance entre la relation R et la relation d'identité afin de justifier la rationalité du souci prudent de Brainy pour Lefty et Righty, Gendler « joue contre son camp », argumentant paradoxalement en faveur de la primauté de la relation R sur la relation d'identité, puisque cette relation R est la structure commune à tous les cas possibles.

### d) Le souci prudent dans un monde parfitien

Gendler affirme que dans un monde parfitien, il n'y aurait pas de souci prudent. Il pourrait y avoir, à la place, un concept similaire de souci prudent pour ses continuateurs selon une relation R. Mais encore une fois, en quoi consiste cette similarité ? En un souci prudent pour ses continuateurs selon une relation R. Et qu'est-ce qui permet à Gendler d'affirmer que ce n'est pas de cela que parle Parfit ? Le but de Parfit n'est-il pas de dégager des concepts plus larges qu'à la fois, l'identité et le souci prudent tel que le comprend Gendler – c'est-à-dire liés inextricablement à la notion d'identité – de manière à englober un certain nombre de cas possibles, et de proposer ces nouveaux concepts comme fondement conceptuel et métaphysique des anciens, et ayant une portée plus large ? C'est faire, semble-t-il, un procès d'intention, que de suggérer que Parfit parle de souci prudent au sens habituel du terme à l'égard d'une situation qui ne serait pas habituelle, elle, alors que Gendler reconnaît pleinement qu'un certain sens du souci prudent – pour ses continuateurs selon une relation R – serait possible.

## e) La méthode de l'accord

Nous passerons à présent à une contre-critique de l'usage fallacieux de la méthode de l'accord que décèle Gendler dans l'argumentation parfitienne au sujet de la fission. Nous montrerons que, contrairement à ce que pense l'auteur, nous pouvons présenter le cas de la fission d'une manière qui laisse les facteurs en jeu indépendants. Ensuite, nous verrons que le cas de la fission est, certes, une explication de valeurs, mais que celle-ci repose d'abord sur une explication causale. C'est avant tout cette explication causale qui est concernée. Puis, sur elle, se greffe une explication de valeurs. Enfin, nous montrerons qu'il y a un sens en lequel le test d'association-dépendance échoue à montrer ce que Gendler veut montrer au sujet du cas de la fission.

Pour Gendler, la combinaison de facteurs dépendants et d'une évaluation portant sur des valeurs peut parfois rendre la méthode de l'accord inutilisable. Le critère que préconise la philosophe pour séparer le bon grain de l'ivraie lorsqu'un tel cas se présente est son test d'association-dépendance. Appliquées au cas de la fission, ces considérations invalideraient les conclusions de Parfit. Rappelons le tableau qui récapitule comment la méthode de l'accord serait censée traiter le cas de la fission :

A = la relation R est obtenue entre X et Y
B = la relation R n'est pas obtenue entre X et quelque non-Y que se soit qui co-existe avec Y
P = le souci prudent de X pour Y est rationnel

| | A | non-A |
|---|---|---|
| B | P | |
| non-B | P | non-P |

Avec ce tableau, on voit que le cas A-plus-B n'est qu'une variante du cas A.

On peut cependant exprimer le cas de la fission d'une manière qui laisse les facteurs en jeu indépendants, ainsi que l'illustre le tableau suivant :

A = la relation R est obtenue entre X et Y (co-existant avec Z)
B = la relation R est obtenue entre X et Z (co-existant avec Y)
P = le souci prudent de X pour Y est rationnel

| | A | non-A |
|---|---|---|
| B | P | non-P |
| non-B | P | non-P |

On considère, bien sûr, que les seuls individus pouvant entretenir une relation R avec X sont Y et Z.

Cette présentation colle mieux, semble-t-il, à l'expérience de pensée présentée par Parfit : selon le succès et/ou l'échec – non prévisible – de l'une et/ou l'autre greffe, Brainy est lié ou non par une relation R à Lefty et il est lié ou non par une relation R à Righty. Ces événements sont indépendants. Ils peuvent se produire l'un sans l'autre ou simultanément.

Les cas A/non-B, non-A/B et non-A/non-B sont assimilables aux cas ordinaires, tandis que le cas A/B est celui décrit par le scénario n°2 de l'expérience de pensée de Parfit.

Le simple fait que nous soyons parvenus à exprimer le cas de la fission d'une manière qui laisse les facteurs en jeu indépendants suffit à installer un gros doute sur la démonstration de Gendler. Toutefois, l'on peut montrer, en outre, que l'évaluation de valeur est, dans le cas de la fission, subordonnée à une évaluation causale. Considérons le tableau suivant :

A = la relation R est obtenue entre X et Y (co-existant avec Z)
B = la relation R est obtenue entre X et Z (co-existant avec Y)
P = X survit (au sens parfitien de ce terme) à travers Y

| | A | non-A |
|---|---|---|
| B | P | non-P |
| non-B | P | non-P |

Il s'agit bien d'une évaluation causale, basée sur des faits. Que X survive ou non à travers Y est, semble-t-il, un fait, et pas simplement une vue de l'esprit. Ce fait est

hypothétiquement causé par une relation R liant X à Y ou X à Z. Ensuite, il suffit à Parfit de montrer qu'il est rationnel de se soucier des « soi » à travers lesquels nous « survivons » pour en déduire le tableau précédent.

Il va sans dire que le test d'association-dépendance n'a plus aucun sens pour ces deux nouveaux tableaux, puisqu'il n'y a plus de cas spécial et de cas général. Nous voudrions toutefois revenir à la manière dont Gendler présente l'application de la méthode de l'accord au cas de la fission pour montrer qu'il y a un certain sens en lequel le test d'association-dépendance échoue à montrer une utilisation fautive de cette méthode dans ce cas.

Gendler soutient que dans un monde parfitien, où fissions et fusions d'individus seraient choses communes, nous n'aurions pas de concept de souci prudent, car ce concept est lié, de manière inextricable, à d'autres concepts, eux-mêmes liés, de manière inextricable, au concept d'identité. Toutefois, elle admet qu'il pourrait exister un concept similaire, de « souci prudent envers les continuateurs selon une relation R ».

Or, de deux choses l'une :

D'une part, tout porte à croire que ce concept n'est pas totalement hétérogène au concept de souci prudent tel que l'entend Gendler. Si ces concepts sont, comme elle le dit elle-même, similaires, c'est qu'ils ont une structure commune, et on se rend compte, en isolant cette structure commune, que l'on dégage ce même second concept de souci prudent pour les continuateurs selon une relation R. Ce dernier concept est donc englobant et plus fondamental que celui de souci prudent ordinaire, lié à l'identité. Dans le cas particulier où la relation R est une relation d'identité, le concept de souci prudent envers les

continuateurs selon une relation R devient un concept de souci prudent ordinaire, tel que le défini Gendler.

D'autre part, rien ne permet de dire que, lorsque Parfit parle lui-même de souci prudent, il n'endosse pas la définition large de souci prudent envers ses continuateurs selon une relation R, plutôt que la définition ordinaire qu'utilise Gendler. Par conséquent, ce n'est sans doute pas trahir la pensée de Parfit que de remplacer, dans l'analyse que Gendler fait de son cas de la fission, la phrase « P = le souci prudent de X pour Y est rationnel » du tableau, par la phrase : « P = le souci prudent (ou un concept apparenté ne présupposant pas le concept d'identité) de X pour Y est rationnel ». Lorsque la relation de X à Y est une relation d'identité, le souci prudent pour les continuateurs selon une relation R devient un souci prudent ordinaire, au sens de Gendler. Néanmoins, si l'analyse de Gendler était juste, concernant le sens que Parfit donne au souci prudent, alors ce serait trahir la pensée du philosophe que de procéder à ce remplacement, mais cela resterait une idée valable, à titre de suggestion, en vue de répondre à la difficulté soulevée par Gendler.

### f) En conclusion

Nous avons d'abord montré que, si Parfit ne montre pas que l'identité ne compte pas en pratique – rationalité pratique – dans les cas ordinaires, il semble pourtant bien qu'il ait montré qu'elle ne compte pas en théorie – rationalité épistémique – à condition d'admettre ses intuitions concernant le cas de la fission.

Ensuite, il découle de cela qu'une approche en facteur commun semble beaucoup plus plausible qu'une approche explicative, au regard des objectifs philosophiques de Parfit.

Concernant la similarité entre les concepts d'identité et de relation R, et également entre les concepts de souci prudent et de souci prudent pour les continuateurs selon une relation R, nous avons montré qu'elle se base sur une structure commune, pour chacune de ces paires de concepts. Cette structure commune, est, respectivement, pour chaque paire, le concept de relation R, et celui de souci prudent pour les continuateurs selon une relation R. Par conséquent, s'il est vrai que le souci prudent de Brainy pour Lefty et Righty est rationnel parce que être lié à quelqu'un par une relation R est très ressemblant au fait de lui être identique, cette similarité repose, ultimement, sur le fait que la relation d'identité est un cas particulier de relation R.

Enfin, nous avons montré que la méthode de l'accord n'était pas du tout mal utilisée par Parfit. Entre autres choses, on peut interpréter son expérience de pensée de manière à ce que les facteurs en jeu soient indépendants.

Nous ne souscrivons pas nécessairement aux vues de Parfit concernant la nature et l'importance de l'identité personnelle. Toutefois, la méthodologie qu'il emploie ne nous apparaît pas foncièrement erronée dans sa démarche. Sans doute peut-on critiquer les conclusions qu'il tire de ses expériences de pensée, mais pas de la manière dont le fait Gendler. Concrètement, la manière dont Parfit évalue les scénarios qu'il présente nous apparaît contestable (cf. ce que nous avons dit de ses expériences de pensée dans cette thèse), mais, à supposer que ces intuitions soient admises, la généralisation aux cas ordinaires à laquelle il procède ne nous semble pas si problématique que ce qu'en pense Gendler.

### *C. Bilan sur Gendler*

Suite à l'analyse de ces deux articles, une interrogation se fait pressante : Gendler se rend-t-elle compte de l'enjeu métaphysique de la question de l'identité personnelle ? Par delà les critiques apportées par elle, au sujet des expériences de pensée impliquant le concept de personne – et auxquels nous avons essayé de répondre – on a l'impression que, pour la philosophe, de toute façon, le but des expériences de pensée, dans le cadre des réflexions sur l'identité personnelle, n'est jamais que de peaufiner un concept, et que les croyances que ces scénarios imaginaires sont censés évoquer, sont des croyances sur nos concepts, leur nature, leur utilisation, la portée de leur validité.

Alors que pour nombre de philosophes, il semble bien que ce que l'on recherche, ce n'est pas tant à mieux circonscrire nos concepts, leur sens conventionnel, et à changer nos croyances les concernant – dans une perspective que l'on pourrait qualifier d'instrumentaliste – que de mieux circonscrire les *choses* auxquelles ces concepts sont supposées se référer parfois, et tout au moins le *sens réel* – et non pas conventionnel – de ces concepts, ainsi que de changer nos croyances sur ces choses et sur ce sens réel – dans une perspective que l'on pourrait qualifier de réaliste.[265]

---

[265] Que dans ses réflexions, Gendler prenne le parti de ramener l'ensemble des discussions sur l'identité personnelle à une simple analyse conceptuelle, sans portée ontologique, nous semble attesté par un certain nombre de ses remarques, qui parsèment ses deux articles :

- Gendler semble considérer que nos concepts – certains au moins, comme celui de personne – ne sont fiables que dans le cadre étroit des usages qui ont contribué à leur formation. Mais cela implique fortement qu'il n'y a pas de réalité extra-cognitive à laquelle ces concepts se réfèrent.

Cette distinction est capitale : y réfléchir revient à se demander si, que l'on soit mort ou vivant demain, torturé ou riche (dans l'expérience de Williams), n'est que pure vue de l'esprit, manière de décrire diversement une même situation, question de préférence personnelle ou de consensus sociolinguistique, etc. ou bien s'il s'agit, au-delà de ces considérations, de parler de choses réelles qui pourraient se passer, d'une façon ou d'une autre selon les cas. Que les évaluations de ces scénarios ne puissent pas être distinguées d'un point de vue empirique ne veut pas dire qu'elles évoquent les mêmes états de choses ou que ces évaluations ne puissent pas être comparées et ordonnées selon un degré de plausibilité. L'empirie n'est pas le seul critère. Même en science, on utilisera le rasoir d'Occam, par exemple, voire des critères esthétiques, afin de faire le tri entre deux descriptions différentes des mêmes faits, ayant exactement les mêmes conséquences empiriques. En philosophie, dans le domaine des réflexions sur l'identité personnelle, les ressentis *subjectifs* peuvent valablement être tenus pour de bons critères discriminants, entre les différentes conceptions de la personne, du *sujet*. Pour justifier cette affirmation, disons que le concept d'identité personnelle n'a de sens et ne pose problème que pour autant qu'il est lié à l'existence de *sujets*, et que le problème de l'identité personnelle consiste

---

- Si un concept est prototypique (comme celui de personne, selon Gendler), alors il semble qu'il ne peut être que conventionnel, et non pas se référer à une réalité extra-cognitive. Il semble difficile de se représenter une réalité qui serait elle-même prototypique, et non pas structurée autour de conditions nécessaires et suffisantes. Ce serait une réalité fluctuante et vague, qui dépendrait de l'état cognitif des sujets. (Notons toutefois qu'il existe une métaphysique du vague et de l'imprécis qui défend précisément cette idée, mais elle n'est pas sans poser un certain nombre de problèmes).

- Par « rationalité du souci prudent » (lors de la critique de l'argument de la fission), Gendler semble n'être intéressée que par une rationalité pratique – celle qui permet d'agir de manière efficace dans le monde – et non pas épistémique – celle qui permet de connaître au mieux le monde tel qu'il est. Elle récidive en affirmant que la lecture en facteur commun n'est pas adéquate aux besoins de Parfit. Elle ne se pose pas la question de savoir si le concept de personne ne pourrait pas avoir, par delà ses usages pratiques, un sens réel.

précisément en celui de l'accord entre nos intuitions et ressentis sur nous-mêmes, la réalité, et le possible métaphysique. Si on décide de faire fi de la subjectivité pour réfléchir à l'identité personnelle, alors cette question n'a plus aucun sens, plus aucun intérêt philosophique. Laissons faire la science ! Que viennent faire les philosophes là-dedans ? Et sans parler des éthiciens qui s'arrogent le droit de contrôler nos actions avec ces vaseux concepts de souffrance, de bonheur, de responsabilité morale, de culpabilité, de libre-arbitre, etc. Si l'on fait fi de la subjectivité, pour parler de l'identité personnelle, alors des situations comme celles-ci sont – semble-t-il – rationnelles :

- « Vous dites que vous avez mal, mais ces indicateurs physiologiques me démontrent que vous ne ressentez pas de douleur, contrairement à ce que vous croyez, alors merci de cesser de vous plaindre et rendez vous à l'évidence ! »

- Votre mère vient de mourir, mais qu'à cela ne tienne, vous la remplacez sans problème, comme n'importe quel objet, par une femme d'à peu près le même âge et disposant d'à peu près les mêmes qualités.

- L'esclavage, la torture, le viol, le meurtre, etc. ne sont pas des crimes.

- Personne n'est responsable de ses actes. Le comportement de chaque individu n'est que le produit d'un déterminisme des gènes et de l'environnement.

- Etc.

Dans le chapitre suivant, nous nous employons à isoler cette réalité métaphysique dont, pensons-nous, les expériences de pensée sur l'identité personnelle s'attachent à dégager les conditions d'identité au cours du temps. Mais avant, demandons-nous quel peut bien être l'utilité d'une telle démarche.

# IV. Métaphysique du « moi » : utilité et fondement.

## *A. A quoi ça sert ?*

A notre époque, il semble que tout, pour avoir un quelconque intérêt, se doit d'être utile, productif, etc. Plus même, que cela soit directement utile, utilisable en pratique, concrètement, dans la vie de tous les jours. Il faut que cela serve notre bonheur et confort matériel au jour le jour. A un sceptique qui doutait de l'intérêt de ses expériences, Benjamin Franklin répliqua, dit-on : « A quoi sert un enfant nouveau-né ? ». On voit bien aussi, de nos jours, les conséquences néfastes d'une politique de recherche uniquement centrée sur les applications. La recherche fondamentale est un creuset nécessaire de jaillissement de nouvelles idées. Même si toutes ces idées ne semblent pas avoir d'applications pratiques sur le moment, elles peuvent, bien plus tard, en avoir d'étonnantes. Mépriser la recherche fondamentale au prétexte qu'elle ne produirait rien d'intéressant, c'est se focaliser sur les petits profits à court terme. Comme le disait Louis Pasteur : « Il n'y a pas de recherche appliquée, il n'y a que des applications de la recherche fondamentale ». Ce n'est peut-être pas entièrement vrai car parfois, c'est au contraire la recherche de solutions à un problème pratique qui peut avoir des répercussions fondamentales. Toutefois, le mouvement inverse est au moins aussi fréquent. L'histoire des sciences nous enseigne que ce qui peut paraître inutile à une époque peut avoir des répercussions technologiques importantes à une autre.

Tout au long de l'histoire de l'humanité, de nombreuses recherches ont abouti à des découvertes et des applications majeures qui n'étaient pas attendues. Ainsi, le laser, qui permet aujourd'hui, entre autres choses, d'écouter de la musique sur CD, de soigner la cataracte et des tumeurs cancéreuses, de découper avec précision des matériaux, etc. n'a pu voir l'existence que grâce à des travaux, de nature purement fondamentale, réalisés par Albert Einstein en 1917, soit plus d'un demi-siècle avant l'invention – grâce à ces travaux fondamentaux – du premier laser !

Nous avons surtout parlé de recherche scientifique jusqu'ici, et nous avons semblé vouloir comparer le travail de la philosophie à celui de la science fondamentale, pour redorer le blason de celle que l'on appelle parfois pompeusement « la reine des sciences ». Ce serait faire de l'imposture philosophique caractérisée. Tel n'est heureusement pas notre but. Si la philosophie est ici comparée à la science fondamentale, il faut bien préciser de quel type de philosophie nous parlons, et à quoi nous l'opposons. C'est d'ontologie ou de métaphysique dont il s'agit, et nous l'opposons à des domaines plus spécialisés et plus proches de l'action et de l'humain concret que sont l'épistémologie, l'éthique, la philosophie politique, l'esthétique, etc. Cela dit, toutes ces disciplines peuvent éventuellement et indirectement avoir des répercussions sur le travail de la science, tant fondamentale qu'appliquée. Pensons par exemple à la gestion des fonds de recherche, qui peut être influencée par des considérations éthiques et politiques. Pensons aussi, actuellement, à la résurgence du mouvement néo-créationniste en Amérique, à travers sa version présentable, la doctrine du « Dessein Intelligent » (*Intelligent Design* en anglais, que nous abrégerons ID). Il faut une bonne critique philosophique des fondements de cette doctrine et une solide compréhension philosophique de la science, pour bien en saisir l'aspect pernicieux et pseudoscientifique. Or, une théorie telle l'ID n'est pas innocente. En confondant les

genres, c'est-à-dire en proposant de remplacer une théorie scientifique, certes, non parfaite, mais testable, par une théorie métaphysique indéboulonnable car blindée contre la réalité, elle est dangereuse pour l'activité scientifique et pour son progrès. En effet, elle sape ce qui en fait le fondement : rechercher des explications logiques basées sur des faits empiriques. Ce faisant, elle tue toute motivation à chercher et donc fait paraître l'activité scientifique inutile. En effet, une fois que l'on a posé Dieu comme explication scientifique valable, alors il n'y a plus rien à chercher : face à tout phénomène inexpliqué, on pourra toujours invoquer Dieu – dont on ne sait pas grand-chose, si ce n'est qu'il pourrait en principe **tout faire** – comme explication valable, et l'investigation sera terminée… Comme on ne pourrait pas imaginer une situation qui pourrait éventuellement infirmer cette explication par le divin, alors on serait garanti qu'elle serait définitive. Le problème est que, pour la même raison, elle serait complètement stérile. Où en serions-nous, aujourd'hui, si face à tous les phénomènes que nous ne comprenions pas, nous avions renoncé à rechercher des causes naturelles vérifiables pour nous contenter des causes surnaturelles invérifiables ? Certainement pas à notre société des ordinateurs et logiciels de traitement de textes, qui permettent paradoxalement aux promoteurs de l'ID de rédiger leur propagande.

Le biologiste Richard Dawkins explique qu'il a un jour déclaré à un philosophe, au cours d'un dîner, qu'il ne pouvait pas imaginer être athée avant 1859, année de la parution de *L'origine des espèces* de Darwin[266]. Ce qui revient implicitement à critiquer l'attitude des athées du XVIIIème siècle. Ce à quoi Jean Bricmont répond très justement :

[266] Voir DAWKINS, Richard. *The Blind Watchmaker*. Bricmont (BRICMONT, Jean. « Science et religion : l'irréductible antagonisme ») note que :

> « [I]maginons, ce qui est évidemment impossible, qu'on démontre demain que toutes les données géologiques, biologiques et autres sur l'évolution sont une gigantesque erreur et que la Terre est vieille de 10 000 ans. Ceci nous ramènerait plus ou moins à la situation du 18e siècle. Nul doute que les croyants, surtout les plus orthodoxes, pousseraient un immense cri de joie. Néanmoins, je ne considérerais nullement cette découverte comme un argument en leur faveur. Cela montrerait que nous n'avons, après tout, pas d'explication de la diversité et de la complexité des espèces. Bien ; et alors ? Le fait que nous n'ayons aucune explication d'un phénomène n'implique nullement qu'une explication qui n'en n'est pas une (par exemple, une explication théologique) devient subitement valable.
>
> La célèbre phrase de Jacques Monod : "L'homme sait enfin qu'il est seul dans l'immensité indifférente de l'Univers d'où il a émergé par hasard"[267] souffre également d'une certaine ambiguïté, qu'on retrouve chez certains biologistes ; que veut dire ici le mot "hasard" ? S'il signifie que l'homme n'était pas prédestiné, ce n'est pas réellement une découverte scientifique ; les explications en terme de causes finales ont été abandonnées pour des raisons similaires à celles qui ont mené à l'abandon des explications de type religieux (impossibilité de les formuler de façon à ce qu'elles soient testées). Mais si le terme désigne ce qui n'a pas de causes (antécédentes), alors la phrase exprime simplement notre ignorance concernant l'origine de la vie ou certains aspects de son évolution. Le hasard n'est pas plus une cause ou une explication que Dieu[…]. En fin de compte, le Dieu soi-disant découvert par la science, comme le hasard, n'est qu'un nom que nous utilisons pour recouvrir notre ignorance d'un peu de dignité. »[268].

Quelle peut bien être l'utilité d'une recherche de critères de l'identité personnelle qui fonctionnent dans tous les cas possibles ? La considération de tout ce qui vient d'être dit pourrait dispenser de répondre à la question et renvoyer à dans un siècle ou deux.

---

> « Dawkins explique correctement l'argument sceptique et pré-darwinien de Hume, mais il ne semble pas apprécier le fait que de tels arguments sont toujours nécessaires, même après Darwin, pour faire face par exemple à l'argument anthropique. La découverte de Darwin déplace le "problème" lié à l'argument basé sur la finalité apparente de l'univers, mais il ne le résout pas. La solution passe, même aujourd'hui, par une critique philosophique de la religion. Ceci dit, il n'y a pas de doute que le darwinisme a apporté un immense soutien psychologique à l'athéisme. »

267 MONOD, Jacques. *Le hasard et la nécessité*.

268 BRICMONT, Jean. « Science et religion : l'irréductible antagonisme ».

Toutefois, nous éprouvons le besoin de ne pas laisser notre lecteur sur sa faim, et d'explorer quelques pistes… Mais bien sûr, ce faisant, nous nous engageons sur le terrain glissant de la spéculation et de la philosophie-fiction.

Voici quelques exemples d'éventuelles applications futures de la recherche d'un critère d'identité personnelle couvrant tous les cas possibles :

**Le télétransport :** On sait désormais téléporter l'état quantique de très petites particules[269]. Peut-être qu'un jour nous serons en mesure de faire de même avec des êtres vivants, et en particulier des êtres humains. Si cela devenait possible, il serait éthiquement essentiel de s'interroger sur l'identité personnelle du voyageur avant et après téléportation. En effet, si cette identité n'était pas la même, cela voudrait dire que le télétransport, loin d'être le moyen de transport le plus rapide, serait la méthode de suicide la plus discrète, la plus propre et la moins génératrice de deuils. Imaginons que la vie me peine et que je veuille en finir, sans pour autant blesser mes proches, je n'aurais qu'à me « téléporter ». L'individu résultant ne serait pas moi mais juste un réplicat indiscernable. Bien sûr, comme ce réplicat serait qualitativement le même que moi et aurait l'impression d'être moi, il pourrait continuer d'avoir des velléités suicidaires. En ce cas, il aurait juste à se « téléporter » à nouveau, et ainsi de suite.

---

269 Selon Wikipedia (<http://fr.wikipedia.org>), à l'article « Téléportation quantique » :

> « La **téléportation quantique** est une technique discutée dans le cadre la théorie quantique de l'information pour transférer un état quantique à travers l'espace, en utilisant des états intriqués et la transmission d'une information *classique*. […] La première démonstration expérimentale de l'intrication fut effectuée à l'université d'Innsbruck (en Autriche) en 1997, par le groupe de Anton Zeilinger. Leur configuration permit de téléporter l'état quantique de polarisation à travers une table optique. Une expérience ayant eu lieu quelque temps plus tôt à Rome, par le groupe de De Martini, avait montré le principe de la téléportation, mais en impliquant seulement deux particules, au lieu de trois : une transportant l'information à téléporter, et deux particules intriquées, qui fournissent le canal quantique et doivent être indépendantes de celle transportant l'état.
>
> En 2004, un autre groupe de recherche à l'université d'Innsbruck et un groupe au NIST démontra la téléportation d'atomes : l'état quantique électronique d'un ion de Calcium fut téléporté vers un autre, l'ensemble des ions étant contenu dans un piège à ions linéaire. »

**Les prothèses cérébrales :** On sait actuellement remplacer de nombreux organes par des prothèses artificielles fonctionnellement équivalentes. Viendra peut-être un jour où nous pourrons remplacer de plus en plus d'organes de plus en plus complexes… et peut-être même le cerveau ! Mais peut-on remplacer le cerveau d'une personne par une prothèse artificielle fonctionnellement équivalente sans tuer de facto cette personne ? Un fonctionnaliste, qui aurait, de plus, *Reasons and Persons* comme livre de chevet, ne verrait certainement pas d'inconvénient majeur à ce type d'opération. Peut-être commettrait-il là la plus grande erreur de sa vie ? Inversement, un partisan du critère cérébral refuserait sans aucun doute une proposition de prothèse cérébrale (ou même de greffe d'un cerveau préalablement configuré de manière à ce que sa configuration synaptique soit semblable à celle du receveur). Ne risquerait-il pas de se priver et de priver ses proches de quelques années supplémentaires ? Le seul moyen de répondre à ce type de questions sera de disposer d'un critère optimal de l'identité personnelle qui couvre ces cas « baroques ».

**L'après-mort :** *« L'immortalité de l'âme,* disait Pascal, *est une chose qui nous importe si fort, et qui nous touche si profondément, qu'il faut avoir perdu tout sentiment pour être dans l'indifférence de savoir ce qui en est. »*[270]. Or, la réflexion sur l'identité personnelle peut sans doute apporter quelques éclairages sur cette question. Mais pour cela, il faut nécessairement l'amener vers des contrées inhabituelles, la confronter à des cas limites et loufoques du genre de ceux qui sont critiqués par les gens qui ne voient dans la question de l'identité personnelle qu'un problème ordinaire pour des utilisations ordinaires. Dans une perspective inversée, certaines personnes peuvent avoir de bonnes raisons, indépendamment de toute considération sur l'identité, de supposer qu'il y a une certaine forme de survie après la mort. Pour elles, il s'agira alors de donner de l'intelligibilité à cette survie, en recherchant un

270 PASCAL, Blaise, *Les Pensées.*

critère de l'identité personnelle qui, tout en étant adéquat pour les cas ordinaires, rende possible une survie après la mort, telle qu'elles la conçoivent. Pour notre part, nous développerons justement cette thématique de la survie, notamment dans les annexes, en jouant à la fois sur ces deux plans. D'une part, nous montrerons qu'une certaine forme de survie semble découler de certains critères plausibles de l'identité personnelle. D'autre part, nous montrerons que la notion de néant est un faux concept. Par conséquent, un critère adéquat de l'identité personnelle devra être compatible avec une certaine forme d'éternité.

Mais enfin, nul n'est besoin de verser dans l'eschatologie ou la futurologie pour trouver, dès maintenant, des applications à la recherche d'un critère de l'identité personnelle applicable à tous les cas possibles. Ces applications se trouvent en éthique.

La question de l'identité personnelle peut se poser dans le cas, par exemple, du statut de l'embryon humain. L'embryon humain est-il une personne ? A partir de quand et sur la base de quels critères peut-on dire qu'il s'agit d'une personne ? Il est important d'apporter une réponse à cette question car, selon que l'on considère l'embryon humain comme une personne ou non, on n'a pas les mêmes droits et devoirs envers lui. Par exemple, on ne peut pas se permettre des expériences qui mettraient sa vie en danger si on considère qu'il s'agit d'une personne. Il faut également refuser l'I.V.G. Quand on a conscience de l'enjeu actuel des questions d'éthique et de bioéthique, en matière de recherche et développement, et quand on prend conscience de la place essentielle que prend la notion de personne, de conscience, de libre-arbitre, de sentiment, dans ces questions, on ne peut pas se dire que la question de l'identité personnelle n'est qu'un joujou philosophique accessoire sans portée concrète. La vision de la personne qui sera retenue aura une répercussion directe sur les principes d'éthique qui seront défendus, lesquels auront un poids considérable sur l'activité

des chercheurs et ingénieurs. Alors que certaines expérimentations seront interdites, d'autres seront au contraire valorisées (par exemple les recherches visant à améliorer la santé et les conditions de vie des gens).

La question de l'identité personnelle se pose également en ce qui concerne la responsabilité morale des personnes. Si quelqu'un, suite à un trouble passager, commet un meurtre, mais que par la suite, il n'a plus souvenir de son acte en tant qu'auteur de cet acte et qu'il porte sur cet acte un jugement négatif, alors a-t-on le droit de le tenir pour responsable et de le punir ? C'est cette question qui avait motivé Locke dans ses interrogations au sujet de l'identité. Dans *L'Essai sur l'Entendement Humain*[271], Locke déclare :

> « Mais au grand et redoutable jour du jugement, où les secrets de tous les cœurs seront découverts, on a droit de croire que personne n'aura à répondre pour ce qui lui est entièrement inconnu et que chacun recevra ce qui lui est dû, étant accusé ou excusé par sa propre conscience »

## *B. Y a-t-il un « moi » réel, avec des conditions d'identité à travers le temps ?*

### 1. Réalisme ou anti-réalisme ?

Concernant la nature profonde des choses, l'en-soi, il existe deux positions extrêmes que sont le réalisme et l'anti-réalisme. Pour le réalisme, les meilleures théories dont nous disposons tendent asymptotiquement à décrire correctement une réalité qui existe indépendamment de notre capacité à la saisir ; ce que nie l'anti-réalisme.

---

271 LOCKE, John. Essai sur l'entendement humain. Livre II, Chapitre XXVII, « De l'identité ».

Plus précisément, le réalisme affirme deux choses :

(1) Il existe un monde objectif, c'est-à-dire un monde indépendant de nos représentations ; c'est une thèse ontologique (le **réalisme ontologique**).

(2) Le monde est correctement décrit ou descriptible par les théories scientifiques, du moins les meilleures que nous pouvons formuler ; bref, le monde est connaissable ; c'est une thèse épistémologique (le **réalisme épistémologique**).

La négation de la première thèse est l'idéalisme (dont une variante extrême est le subjectivisme ou même le solipsisme). La négation de la seconde est l'instrumentalisme proprement dit. Le contraire de l'idéalisme est plus particulièrement le réalisme ontologique. Le contraire de l'instrumentalisme est plus particulièrement le réalisme épistémologique.

Pour l'idéalisme, il n'y a pas de monde indépendant de nos systèmes de représentation. Un argument en faveur de l'idéalisme est fourni par le rasoir d'Occam : puisqu'un hypothétique monde indépendant de nos représentations ne peut être connu qu'au moyen de nos systèmes de représentation, ce qui est connu de ce monde à un moment donné, est dépendant de nos systèmes de représentation. Ce qui est inconnu de ce monde, à ce moment, ne peut pas, par définition, avoir la moindre influence sur nos systèmes de représentation, car alors il serait connu. Par conséquent, tout se passe comme si, ce qui est inconnu de nos systèmes de représentation à un certain moment – c'est-à-dire,

ce qui est indépendant de nos systèmes de représentation à ce moment – n'existe tout simplement pas. Le rasoir d'Occam nous dit de ne pas multiplier les entités sans nécessité. Par conséquent, ce que l'on ne connaît pas du monde – c'est-à-dire ce qui est indépendant de nos représentations – n'existe pas, tant qu'on ne le connaît pas (ce qui rejoint la conclusion de Berkeley, lui aussi idéaliste : « Etre, c'est être perçu ou percevoir »). De là, l'idée, promue par un certain constructivisme radical, suivant laquelle nous construisons littéralement le monde par nos représentations. L'idéalisme n'est pas forcément instrumentaliste, bien au contraire. Ainsi, si le monde se réduit aux représentations qu'on en a, alors il en découle assez naturellement que le monde est connaissable, car tout ce que l'on croit savoir sur le monde est identique à ce que l'on sait sur le monde.

Selon l'instrumentalisme, un monde indépendant de nos représentations peut bien exister, mais les théories qui sont supposées le décrire ne sont que des fictions utiles produites par notre système cognitif dans un but fonctionnel. On ne doit pas attendre d'elles qu'elles correspondent à la réalité. Un argument en faveur de l'instrumentalisme est que, pour savoir si une proposition est vraie, il faut savoir si elle correspond à la réalité, mais que, pour savoir si elle correspond à la réalité, il faut connaître la réalité. Un autre argument peut être tiré du néodarwinisme. Pour le réalisme, le néodarwinisme est vrai. Or, le néodarwinisme nous apprend que notre cerveau, qui nous permet d'acquérir des croyances, est un produit contingent de la sélection naturelle. Nos croyances sur le monde sont donc garanties quant à leur valeur adaptative – survie, reproduction – mais pas quant à leur valeur épistémique – leur véracité. Au contraire, il serait même très étonnant que l'évolution nous ait doté d'un cerveau capable de saisir l'ensemble du réel. C'est une possibilité très improbable, affaiblie encore si on considère les autres espèces vivantes qui n'en sont pas capables, encore moins que nous. Prétendre le contraire serait nous attribuer

une place à part dans le monde vivant, ce que peu de réalistes accepteront, car aucune donnée scientifique ne nous incite à le penser. Par conséquent, si le réalisme est correct, le darwinisme est correct. Mais si le darwinisme est correct, alors le réalisme est faux (très probablement). Donc le réalisme est faux (très probablement).

En caricaturant, pour l'idéalisme, toute proposition construite selon certaines règles et cohérente avec l'ensemble du système de propositions jugées vraies, est vraie. C'est une conception **cohérentiste** de la vérité. Pour l'instrumentalisme, aucune proposition n'est vraie, ou alors elle est vraie par hasard. Pour le réalisme, enfin, toute proposition correspondant au réel est vraie. C'est une conception **correspondantiste** de la vérité.

Le réalisme et l'anti-réalisme (idéalisme ou instrumentalisme) peuvent être des options ontologiques globales ou limitées à certaines catégories d'êtres. Ainsi, il est fréquent d'être réaliste au sujet des objets physiques, et anti-réaliste au sujet des objets mathématiques. Les platoniciens adoptent le choix inverse. Un choix ontologique n'est jamais entièrement justifiable, mais on peut avoir de bons motifs qui nous font pencher pour l'un plutôt que l'autre.

En ce qui nous concerne, il nous apparaît, en général plus raisonnable de s'en tenir à un « réalisme modéré », une saine suspension de jugement quant à l'existence d'un monde indépendant de nos représentations, et quant à la possibilité d'accéder, par nos meilleures théories scientifiques, à la réalité ultime. Si notre système cognitif est un produit de l'évolution, il est hautement peu probable qu'il nous permette d'accéder à la réalité ultime des choses. Le réalisme dur pêche par orgueil. Inversement, affirmer que nos théories scientifiques ne sont que des fictions utiles implique que cette même théorie qui soutient

cela est aussi une fiction utile. L'instrumentalisme se réfute lui-même. Avec l'instrumentalisme, qui plus est, on ne comprend pas bien par quel miracle les théories scientifiques parviennent : à nous conduire vers de nouvelles découvertes, à prédire avec une grande précision les phénomènes, à générer des technologies dont tout le monde peut constater les bienfaits pour peu qu'on sache les utiliser consciencieusement. Enfin, l'idéalisme, ne rend pas justice au fait que nos découvertes prennent souvent à revers nos attentes théoriques, et semblent donc ne dépendre que faiblement de nos représentations. On ne rend pas compte, aussi, de l'intuition commune selon laquelle le monde est quelque chose que l'on découvre et non pas quelque chose que l'on construit.

Comme le disent Alan Sokal et Jean Bricmont[272] :

> « L'objectif de la science est la connaissance de la nature véritable des choses. Cet objectif est ambitieux, mais il n'est pas inaccessible, du moins pour certaines sections de la réalité et si l'on accepte un certain degré d'approximation. »

Celui qui affirme que l'on ne peut rien savoir pense savoir au moins une chose : que l'on ne peut rien savoir. Celui qui affirme que l'on peut tout savoir considère notre raison comme absolue, mais comment le sait-il ? Il donne alors raison à celui qui pense que l'on peut tout savoir parce que le monde se réduit à ce que l'on en connaît, mais comment le sait-il ?

---

[272] SOKAL, Alan et BRICMONT, Jean. *Pseudosciences et postmodernisme. Adversaires ou compagnons de route ?*, p.194.

## 2. Le réalisme du « moi »

Si nous revenons à l'identité personnelle, comment doit-on la situer dans le débat réalisme / idéalisme / instrumentalisme ?

Notons déjà que, un peu paradoxalement peut-être, l'idéalisme, dans sa variante forte qu'est le solipsisme[273], présuppose un réalisme fort du sujet (c'est la moindre des choses…) et donc de l'identité personnelle. On ne peut donc pas facilement contester nos réflexions sur l'identité sur la base de ces variantes de l'idéalisme. Maintenant, qu'en est-il d'un idéalisme objectif qui identifie la réalité, non pas avec les systèmes de représentations de sujets isolés, mais avec le système de représentation généré par l'ensemble des sujets ? (Le

---

273 On distinguera le solipsisme radical ou absolu, pour qui, seul un sujet (moi) existe, position difficilement tenable en pratique et aux conséquences éthiques désastreuses ; du solipsisme convivial (décrit par Hervé Zwirn dans ZWIRN, Hervé. « Mécanique quantique et connaissance du réel ». *In :* D'ESPAGNAT, Bernard (Ed.). *Implications philosophiques de la science contemporaine*) pour lequel autrui peut bien exister mais où chacun vit dans son monde :

> « Du point de vue de la réalité profonde (et non de l'apparence de cette réalité pour nous), seule une fonction d'ondes globale superposée "existe". Dans cette fonction d'ondes, les différents résultats possibles de mesure sont présents et sont corrélés ainsi que tous les états correspondants de l'observateur. La décohérence intervient [...]. Le solipsisme convivial consiste alors à considérer que la conscience de l'observateur est "accrochée" à l'une des branches de la fonction d'ondes ne lui permettant d'observer que la partie classique correspondante. La conscience joue en quelque sorte le rôle d'un filtre ne permettant de voir qu'une partie de la fonction d'ondes globale. Une définition précise de ce processus permet de montrer que les prédictions habituelles de la mécanique quantique sont respectées malgré le fait que la fonction d'ondes n'est jamais rigoureusement réduite. Le point surprenant est alors que rien n'oblige deux observateurs différents à être accrochés à la même branche. Pour une mesure donnée, un observateur peut être accroché à la branche donnant le résultat A alors qu'un autre le sera à la branche donnant le résultat B. Comment peut-il en être ainsi alors qu'on sait que deux observateurs de la même expérience sont — en général — d'accord sur le résultat ? La raison en est que la communication entre observateurs est elle-même un processus de mesure et que le mécanisme d'accrochage garantit la cohérence des observations pour un observateur. Supposons qu'André a observé le résultat A et Bernard le résultat B. Les deux observations ne sont que l'accrochage de la conscience d'André et de Bernard à leur branche propre de la fonction d'ondes globale qui contient les deux possibilités. Si André demande à Bernard ce qu'il a vu, l'interaction entre André et Bernard qui en résulte contient la totalité des possibilités, donc à la fois une branche où Bernard répond A et une branche où Bernard répond B. La fonction d'ondes d'André sera après l'interaction avec Bernard dans un état enchevêtré contenant les deux réponses mais la conscience d'André s'accrochera à la branche correspondant à la réponse cohérente avec son observation précédente, il entendra donc Bernard répondre A conformément à son attente. C'est la raison pour laquelle cette interprétation porte le nom de solipsisme convivial : chaque observateur vit dans son monde qui peut être totalement différent de celui des autres, mais il n'existe aucun moyen de se rendre compte des désaccords et les observateurs sont en parfait accord. Ceci fournit une nouvelle explication de l'intersubjectivité : il n'y a aucun moyen de constater un désaccord. Signalons pour terminer une conséquence étrange sur l'indéterminisme de la mécanique quantique. La fonction d'ondes de l'Univers évolue de manière parfaitement déterministe par l'équation de Schrödinger, seul le mécanisme d'accrochage tire au sort la branche à laquelle chaque observateur s'accroche. Ce n'est donc plus Dieu qui joue aux dés, c'est l'homme, mais avec le constat étrange que deux joueurs peuvent voir le même dé tomber sur une face différente. »

monde 3 de Karl Popper[274]). En fait, une telle position paraît plutôt instable et tend à évoluer, logiquement, vers un réalisme ou au contraire un subjectivisme voire solipsisme. En effet, comment justifier d'une cohérence des représentations de tous les sujets sans poser une réalité qui les transcende tous ? On peut bien vouloir, tel Berkeley, poser Dieu comme sujet suprême qui coordonne les représentations de tous les autres sujets. Mais alors, Dieu, dans ce scénario, joue ni plus ni moins le même rôle que la réalité en soi du réalisme. On pourrait aussi imaginer une sorte d'idéalisme « démocratique » où le réel serait une construction collective en quelque sorte décidée à l'unanimité. Une telle position est proche, semble-t-il, d'un relativisme à la Bruno Latour[275] pour qui la science ne serait qu'une construction sociale dont les résultats ne seraient pas fixés par un réel indépendant exerçant ses contraintes sur l'activité des chercheurs. Cette conception est curieuse. Selon elle, ce serait toujours celui qui crie le plus fort qui aurait raison ou l'opinion la plus répandue qui serait véridique. On devrait donc croire que la Terre était réellement plate jusqu'en 1522, lorsque l'expédition initiée par Magellan autour du globe pris fin. De même, l'Amérique n'existait pas jusqu'à ce que Christophe Colomb la découvre en 1492. On se demande même, alors, comment, dans ces conditions, de telles découvertes ont pu être réalisées. Si la réalité dépend d'un « vote démocratique », alors des « idées marginales » comme celles de nos découvreurs n'auraient jamais pu se trouver validées, à moins qu'elles ne subissent préalablement une percée dans l'opinion publique. Une position idéaliste cohérente en arrive donc au subjectivisme voire au solipsisme. Le subjectivisme, toutefois, souffre d'une autre incohérence. Pour le subjectivisme, ce qui est vrai est ce que le sujet tient pour vrai. Mais devant la multiplicité des croyances contradictoires selon que l'on considère un sujet ou un autre, on se retrouve dans la situation curieuse d'une pluralité de

---

274 Cf. par ex.: POPPER, Karl Raimund, et ECCLES, John Carew, *The Self and its Brain*.

275 LATOUR, Bruno et WOOLGAR, Steve. *La Vie de laboratoire : la Production des faits scientifiques*

vérités incompatibles et donc d'un réel intrinsèquement contradictoire. Pour mettre de la clarté dans cette situation, l'idéaliste, après être passé au subjectivisme, doit à présent devenir solipsiste.

Pour l'instrumentalisme, le critère de validité d'une théorie n'est pas sa vérité mais son utilité. Mais qu'est-ce que l'utilité ? Peut-on définir une utilité **réelle** dans un monde sans sujets **réels** ? Etre utile, c'est toujours être utile pour quelqu'un. Il apparaît donc difficile également, dans la perspective de l'instrumentalisme, de remettre en question la notion de personne. On pourrait objecter à ces deux réflexions, qu'il est peut-être possible de décrire impersonnellement l'utilité. Ainsi, pour John Stuart Mill, fondateur de l'utilitarisme, l'utilité se définit comme la quantité de bonheur totale engendrée[276]. Les conséquences « répugnantes » d'une telle définition sont bien connues : il serait ainsi utile de tuer les gens qui sont plus malheureux qu'heureux ou même de tuer quelqu'un d'heureux, si cela pouvait augmenter la quantité de bonheur totale. Mais même si on doit ne tenir compte que du bonheur total, il n'en demeure pas moins vrai que le bonheur ou le malheur ne flottent pas dans les airs… ils sont vécus, en première personne, par des sujets.

Venons-en à présent au réalisme. Dans un réalisme global, il va de soi que le concept de personne renvoie à quelque chose de réel. Dans un réalisme partiel, on pourrait admettre, en première approximation, que certains concepts correspondent à quelque chose de réel, par exemple les concepts naturels, tandis que d'autres seraient de pures conventions, par exemple les concepts normatifs, ou que d'autres renverraient à une réalité dépendante du sujet, par exemple les concepts psychologiques. On le voit, ce qui ne

---

276 MILL, John Stuart. *L'Utilitarisme*.

relèverait pas du réalisme, relèverait de l'idéalisme ou de l'instrumentalisme. Mais nous avons vu, juste avant, que ces deux théories présupposent un réalisme pour le sujet.

Que l'on adopte donc une attitude réaliste, idéaliste ou instrumentaliste, au fond, cela importe peu relativement au problème de l'identité personnelle, car dans tous les cas, il s'en trouve légitimé, la notion de sujet apparaissant comme quelque chose de réel.

Il reste à explorer notre quatrième option : la suspension du jugement préconisée par le réalisme modéré. Cette conception, et elle seule, pourrait justifier, non pas un anti-réalisme de l'identité personnelle, mais au moins un scepticisme légitime quant à l'identité personnelle. Mais là où le scepticisme est de rigueur, on peut toujours, malgré tout, choisir d'adopter, en connaissance de cause, une option ou une autre, alors pourquoi pas celle de la réalité du sujet ?[277]

Certes, on le peut, mais il reste que l'on peut aussi considérer cela vain, pour les mêmes raisons. Et notre but n'est point de montrer seulement que l'on peut se poser des questions sur les implications du télétransport, de la commissurotomie cérébrale et autre

---

277 Cela pourrait, de plus, être justifié à titre d'idéal régulateur. Le réalisme modéré, bien que n'affirmant pas explicitement l'existence d'un monde indépendant de nos représentations, l'admet généralement sans peine. C'est une option ontologique intuitive, qui donne un sens et un objectif à notre quête de connaissance, même si cette dernière devait rester incomplète et approximative, par limitation de nos capacités cognitives. Toutefois, il existe une marge entre admettre l'existence d'un monde indépendant de nos représentations, et admettre que le sujet en fait partie. Néanmoins, l'existence réelle du sujet est également une option ontologique intuitive que l'on peut accepter comme idéal régulateur. Peut-être pas de la démarche et de la connaissance scientifiques, mais, par exemple, de nos actions, de notre morale. Nous l'avons vu dans le chapitre 2, « Les intuitions antiréductionnistes sur l'identité personnelle », de la partie 2, « Discussion et critique ».

syndrome de personnalité multiple quant à l'identité personnelle, mais aussi qu'un tel questionnement est légitime.

Dans le sous-chapitre suivant, nous répondrons par l'affirmative, en cherchant à mettre en exergue ce « moi » réel dont nous avons suggéré l'existence.

### *C. Quel est donc ce « moi » dont on recherche les conditions d'identité à travers le temps ?*

Wittgenstein, s'inspirant d'une expérience de pensée de John Locke, écrivait :

> « Imaginons un homme qui, les jours pairs, ne se souvient que des événements des jours pairs de sa vie, oblitérant entièrement ce qui s'est passé les jours impairs. Par ailleurs, les jours impairs, il se souvient de ce qui s'est passé les jours impairs précédents ; mais sa mémoire oblitère alors les jours pairs, sans qu'il y ait un sentiment de discontinuité. Si nous le désirons, nous pouvons également admettre qu'il présente une apparence et des caractéristiques différentes les jours pairs et les jours impairs. Devons-nous pour autant dire que deux personnes habitent le même corps ? A savoir, est-il juste de dire qu'il y en a deux et faux de le refuser, ou *vice versa* ? Ni l'un ni l'autre. Car l'usage *ordinaire* du mot "personne" est ce qu'on pourrait appeler un usage composite, convenant aux circonstances ordinaires. »[278].

Autrement dit, la question de l'identité personnelle, appréhendée à l'aide de scénarios fictionnels, n'est pas simplement indécidable suite à quelque limitation cognitive inhérente à l'esprit humain ou à nos moyens conceptuels actuels. Elle est indécidable parce que, d'une certaine manière, elle n'a pas de sens. Le concept de personne ne s'applique pas

---

[278] WITTGENSTEIN, Ludwig. *Le Cahier Bleu et le Cahier Brun*, p. 118-119

ordinairement à des situations extraordinaires. Bref, autant se demander si un couteau sans lame auquel il manque le manche demeure un couteau…

Pourquoi donc cette obstination de certains théoriciens de l'identité personnelle à rechercher un critère de l'identité personnelle qui satisfasse, non pas seulement les cas ordinaires de la vie de tous les jours (et des critères qui fonctionnent dans ces cas là, il faut le dire, on en a déjà), mais aussi tout l'arsenal des cas *borderline* sortis de l'imagination fertile des philosophes et des romanciers de science-fiction ?

Une première réponse, simple, retourne l'argument de Wittgenstein contre lui-même.

Le concept de personne n'a peut-être pleinement de sens que dans le cadre d'usages ordinaires, mais son sens ne s'annihile pas d'un seul coup lorsque l'on passe dans le domaine des *puzzling cases*. Il semble légitime de considérer tout un domaine de cas limites où le concept de personne perd un peu de sa signification, sans toutefois ne plus en avoir du tout. Dirions-nous d'un individu avec les hémisphères cérébraux séparés qu'il n'est plus **du tout** une personne ? Si oui, cela voudrait dire qu'il ne dispose plus d'aucun droit et devoir, qu'on peut lui faire subir ce que l'on veut sans être pour autant immoral… Difficile à admettre !

Le concept ordinaire de personne comporte intuitivement l'idée selon laquelle il y a **toujours** une réponse claire du type « oui » ou « non » à la question de l'identité d'une personne. Par conséquent, c'est le concept ordinaire de personne, et non un concept modifié pour les besoins d'un philosophe, qui implique que la question de l'identité personnelle peut légitimement se poser dans **tous les cas possibles**. Les concepts de

personne qui ne présupposent pas qu'en principe toute question sur l'identité personnelle a une réponse claire en tout-ou-rien (c'est par exemple le cas chez Derek Parfit) ne sont pas ordinaires. Ce sont donc ces concepts pour lesquels l'argument de Wittgenstein a une certaine pertinence.

Une deuxième réponse se trouve dans la suite du passage cité de Wittgenstein :

> « Si j'admets, comme je le fais, que ces circonstances ont changé, l'emploi du terme "personne" ou "personnalité" a par conséquent changé. Et si je souhaite conserver ce terme et lui donner un usage analogue au précédent, je suis libre de choisir entre plusieurs usages, plusieurs genres d'analogies. On pourrait dire dans un tel cas que le terme "personnalité" possède plus d'un héritier légitime »[279].

Certes, le concept de personne inhérent à la littérature des *puzzling cases* n'a sans doute guère à voir avec celui qui est en usage ordinairement, tout au moins lorsque l'analyse de ce concept ne se borne pas à sa dissolution. Sans doute serait-il bon, alors, de préciser clairement ce que l'on entend par *personne*, à chaque fois qu'il est fait usage de ce terme de manière non conventionnelle, tout au plus analogique. Sans doute serait-il meilleur encore d'inventer un nouveau terme pour chaque usage.

Que peut bien être ce sens plus profond qui se cache derrière le concept de personne et qui pourrait justifier de s'appesantir sur les expériences de pensée de double transplantation de cerveau, de personnalités multiples et autres fusions d'individus ? Celui du **soi témoin**, encore appelé soi transcendantal[280] ou soi pur. Que doit-on entendre par

---

279 WITTGENSTEIN, Ludwig. *Le Cahier Bleu et le Cahier Brun*, p. 119.

280 Par Kant (KANT, Emmanuel. *Critique de la Raison Pure*).

là ? Le sujet est délicat à aborder. Il a généré toute une littérature lourde, complexe, et riche en polémiques. Nous voulons parler de la phénoménologie, et de ce qui gravite autour, notamment les deux précurseurs que sont Descartes et Kant. Les critiques, comme nous l'avons vu, ne manquèrent pas, de la part du courant analytique, de l'empirisme logique, de Wittgenstein et de ses successeurs. L'erreur fut probablement de s'enfermer dans une subjectivité auto-fondationnelle, de partir d'une attitude de doute hyperbolique ou d'*époché* peu naturelle. La démarche cartésienne et phénoménologique commet l'erreur de nier le corps, le monde, autrui, puis de faire des contorsions intellectuelles pour essayer de les réintégrer dans leur vision. Mais par leurs prétentions à se constituer en méthodes objectives, à décrire le vécu intime, elles présupposent l'existence d'une communauté linguistique à travers laquelle leurs descriptions peuvent seulement prendre sens. Toutefois, comme le montre bien l'argument du langage privé, ce qui peut être décrit ne peut être que public, d'où l'échec de l'entreprise. Puisque l'on va s'aventurer hors des sentiers battus du sens commun populaire, il importe donc de correctement baliser notre expédition, de crainte de se fourvoyer, à la manière de ce qui est arrivé aux phénoménologues et autres néo-cartésiens et néo-kantiens.

Pour approcher le soi témoin, nous partirons donc de l'univers objectif, et nous y resterons autant que possible. Nous procéderons aussi à notre mise entre parenthèses du monde, mais elle sera, en un sens, objective. Nous ne mettrons pas en doute l'existence de cet univers, nous le viderons – en imagination, par l'isolation sensorielle ou par la méditation – de tout ce qu'il contient. Cela nous ramène singulièrement à la première étape de la critique de la notion de néant par Bergson[281] :

[281] BERGSON, Henri. *L'évolution créatrice*, Chap. IV.

> « Je vais fermer les yeux, boucher mes oreilles, éteindre une à une les sensations qui m'arrivent du monde extérieur : voilà qui est fait, toutes mes perceptions s'évanouissent, l'univers matériel s'abîme pour moi dans le silence et dans la nuit. Je subsiste cependant, et ne puis m'empêcher de subsister. Je suis encore là, avec les sensations organiques qui m'arrivent de la périphérie et de l'intérieur de mon corps, avec les souvenirs que me laissent mes perceptions passées, avec l'impression même, bien positive et bien pleine, du vide que je viens de faire autour de moi. Comment supprimer tout cela ? comment s'éliminer soi-même ? Je puis, à la rigueur, écarter mes souvenirs et oublier jusqu'à mon passé immédiat ; je conserve du moins la conscience que j'ai de mon présent réduit à sa plus extrême pauvreté, c'est-à-dire de l'état actuel de mon corps. Je vais essayer cependant d'en finir avec cette conscience elle-même. J'atténuerai de plus en plus les sensations que mon corps m'envoie : les voici tout près de s'éteindre ; elles s'éteignent, elles disparaissent dans la nuit où se sont déjà perdues toutes choses. Mais non ! à l'instant même où ma conscience s'éteint, une autre conscience s'allume ; - ou plutôt elle s'était allumée déjà, elle avait surgi l'instant d'auparavant pour assister à la disparition de la première. Car la première ne pouvait disparaître que pour une autre et vis-à-vis d'une autre. »

Sur un principe similaire, Ramana Maharshi a développé une méthode de recherche du soi témoin. Dans l'introduction de *Qui suis-je ?*, un ouvrage collectant une série de questions (posées par Shrî Pillai) et de réponses (données par Ramana Maharshi) au sujet de la recherche du Soi, T.M.P. Mahadevan[282] écrit :

> « Le mental est constitué de pensées. La pensée "je" est la première qui s'élève dans le mental. Si l'on poursuit l'investigation "Qui suis-je ?" d'une manière constante, toutes les autres pensées sont détruites, et finalement la pensée "je" elle-même disparaît, laissant la place au Soi non-duel ; les fausses identifications du Soi avec les manifestations du non-Soi, tels que le corps et le mental, cessent, et l'illumination (le *sâkshâtkâra*) s'ensuit. Le processus de la recherche du Soi n'est en aucun cas facile. En posant la question "Qui suis-je ?", de nombreuses autres

[282] MAHADEVAN, T. M. P. *Qui suis-je ?*

> pensées vont surgir; mais, au lieu de leur céder et de les suivre, il faut demander : "A qui se présentent-elles ?". Pour ce faire on doit rester extrêmement vigilant. Grâce à cette investigation constante le mental s'établira dans sa source et ne pourra se disperser et se perdre dans le labyrinthe des pensées créées par lui-même. Toutes les autres disciplines, tels que le contrôle de la respiration ou la méditation sur une image de Dieu, doivent être considérées comme des pratiques accessoires. Elles ne servent qu'à la maîtrise et la concentration du mental. Pour un mental exercé dans la concentration, la recherche du Soi devient comparativement facile. C'est par une investigation incessante que les pensées seront détruites et le Soi réalisé – la Réalité plénière dans laquelle il n'y a même plus la pensée "je", expérience qui est désignée comme "Silence". »

Plus loin, Ramana Maharshi déclare :

> « Je ne suis pas ce corps physique, constitué des sept éléments subtils (*dhâtu*), ni les cinq organes de perception sensoriels, c'est-à-dire l'oreille, l'œil, la langue, le nez et la peau, et leurs fonctions correspondantes : l'ouïe, la vue, le goût, l'odorat et le toucher ; je ne suis pas les cinq organes d'activité, c'est-à-dire les organes vocaux, les mains et les pieds, l'organe de procréation et l'anus, et leurs fonctions respectives : le langage, les mouvements du corps physique, la jouissance et l'excrétion ; je ne suis pas les cinq forces vitales, le *prâna* […], etc. qui permettent d'accomplir leurs fonctions correspondantes ; même l'esprit pensant je ne le suis pas ; et pas non plus cet état d'ignorance inconsciente, dans lequel ne se trouvent que les impressions des objets, et non les objets eux-mêmes et leurs fonctions. […] Après avoir rejeté tout ce qui a été mentionné ci-dessus comme n'étant "pas ceci ni cela", cette pure Conscience qui seule demeure - CELA je suis. »

Avec une méthode comparable, écartons en pensée, l'un après l'autre, les éléments de l'univers. Supposons qu'il ne reste plus, dans ce dernier, outre nous, qu'une balle de golf luminescente que nous contemplons, disons, à une distance d'un mètre. Pourtant nous ne voyons pas qu'une balle de golf, nous voyons aussi tout un espace noir et vide autour.

Supprimons cette balle de golf, l'espace vide et noir reste. Peut-on supprimer cet espace ? Mais par rapport à quoi cet espace disparaîtrait-il ? Qu'y a-t-il autour de lui qui constituerait un plus ultime fond sur lequel son existence – dès lors contingente – se détacherait ? On ne peut parler d'apparitions, de changements et de disparitions que relativement à certain référentiel. On ne peut pas parler d'apparitions, de changements ou de disparitions du référentiel lui-même qui sert à en parler. Ainsi, lorsque nous croyons avoir supprimé tout, il reste encore quelque chose, ce dans quoi tout le reste se déploie. On ne peut supprimer ce résidu, ce grand tableau noir sur lequel, semble-t-il, se dessinent les phénomènes observables.

Comme le note Bergson[283] :

> « Admettons pourtant que le vide absolu soit possible ; ce n'est pas à ce vide que je pense quand je dis que l'objet, une fois aboli, laisse sa place inoccupée, car il s'agit par hypothèse d'une *place,* c'est-à-dire d'un vide limité par des contours précis, c'est-à-dire d'une espèce de *chose*. Le vide dont je parle n'est donc, au fond, que l'absence de tel objet déterminé, lequel était ici d'abord, se trouve maintenant ailleurs et, en tant qu'il n'est plus à son ancien lieu, laisse derrière lui, pour ainsi dire, le vide de lui-même. Un être qui ne serait pas doué de mémoire ou de prévision ne prononcerait jamais ici les mots de « vide » ou de « néant » ; il exprimerait simplement ce qui est et ce qu'il perçoit ; or, ce qui est et ce qu'on perçoit, c'est la *présence* d'une chose ou d'une autre, jamais *l'absence* de quoi que ce soit. »

Quelle est la nature de ce résidu ? Manifestement, il s'agit de l'espace-temps lui-même. En effet, il semble bien que ce soit le référentiel ultime – pour nous – où

283 BERGSON, Henri. *L'évolution créatrice*, Chap. IV.

s'inscrivent les phénomènes observés[284]. Il les présuppose tous (aucun fait n'existe qui ne soit spatio-temporellement situé) et il ne peut pas apparaître, changer ou être détruit[285]. Est-il ontologiquement différent des phénomènes dont il constitue la scène ? Manifestement, non. On peut se représenter ce résidu vide de tout contenu, ou du moins de la plupart des contenus – on admettra que la noirceur de l'espace n'est pas forcément une absence de propriété, et qu'au contraire, une blancheur immaculée pourrait peut-être recevoir le même statut. On ne peut pas, par contre, se représenter un quelconque contenu, phénomène, ou fait, sans *ipso facto* le situer dans un certain espace et une certaine durée. A quoi pourrait bien ressembler une note de musique sans durée ? Une chaise de dimension nulle ? Tout fait, tout phénomène, contient donc en lui-même la spatio-temporalité dont nous avons la forme pure à travers le « silence éternel de ces espaces infinis » qui effrayait tant Blaise Pascal.

Revenons au **soi témoin**. Le soi témoin n'est, en première approximation, rien d'autre que cet espace-temps épuré qui fait nécessairement partie de nous, puisqu'il est le point commun à toute réalité, l'ensemble vide qui est membre de tous les ensembles.

---

[284] Pour Kant (KANT, Emmanuel. *Critique de la Raison Pure*), l'espace et le temps sont des formes a priori de la perception. Il faut comprendre par là que l'espace et le temps sont donnés dans toute perception, qu'ils constituent une condition transcendantale nécessaire à la perception.

[285] Si un début, un changement ou une fin de l'espace est difficile à concevoir – à moins d'imaginer que notre espace existe dans un hyperespace plus vaste – cela est encore plus saillant concernant le temps : comment le temps pourrait-il cesser, puisqu'une cessation implique un après ? Comment pourrait-il commencer, puisqu'un commencement implique un avant ? Comment pourrait-il changer, puisqu'un changement implique une vitesse de changement, et que cette vitesse se mesure par rapport au temps ?

Il est donc universel et n'a rien de particulier. C'est un peu le **point de vue de nulle part** de Thomas Nagel. Nous pouvons nous demander le rapport entretenu avec l'identité personnelle.

Wittgenstein, nous semble-t-il, apporte déjà quelques éléments de réponse à cette question dans son *Tractatus*[286] :

> « 5.641 – Il y a donc réellement un sens selon lequel il peut être question en philosophie d'un je, non psychologiquement.
>
> Le je fait son entrée dans la philosophie grâce à ceci : que "le monde est mon monde".
>
> Le je philosophique n'est ni l'être humain, ni le corps humain, ni l'âme humaine dont s'occupe la psychologie, mais c'est le sujet métaphysique, qui est frontière – et non partie – du monde. »

Le soi témoin (le « je philosophique ») est l'espace-temps épuré (le « sujet métaphysique ») parce qu'il n'est pas un phénomène, mais ce sur quoi se détachent les phénomènes (il est « frontière – et non partie – du monde »).

Voyons cela en détail. Pour commencer, il nous faut remplir à nouveau ce vide, très progressivement. Mettons que nous le remplissions avec une balle de golf luminescente **bleue**. Nous avons donc à présent une balle de golf luminescente bleue qui se détache sur le fond noir de l'espace. Ajoutons à présent une deuxième balle de golf – luminescente également – **rouge**.

Cette balle se trouve dans le même référentiel que la précédente. C'est le même vide – numériquement identique – qu'elle remplit (même si, bien sûr, elle n'en remplit pas la

---

286 WITTGENSTEIN, Ludwig. *Tractatus Logico-Philosophicus*, § 5.641.

même partie au même moment). On voit mal comment il pourrait en être autrement. Rappelons que ce référentiel, par hypothèse, est ultime. Il est donc unique. Si l'on devait imaginer un autre référentiel, on pourrait le comparer au premier, et les deux seraient donc inclus dans un troisième, plus vaste, plus ultime. La seule solution, pour que les deux balles ne se trouvent pas dans le même référentiel simultanément, serait donc d'en supprimer purement et simplement une des deux. Mais alors il n'y aurait plus qu'une balle. On se retrouverait dans la situation précédente.

Les deux balles se trouvant donc dans le même référentiel, il s'ensuit nécessairement qu'elles se trouvent sous un certain rapport l'une envers l'autre. Elles vont, par exemple, se trouver plus ou moins éloignées l'une de l'autre.

Jusque maintenant, nous n'avons exposé que des propositions factuelles. Il s'agissait de balles de golf réelles, situées dans l'espace-temps physique. Il nous faut à présent introduire la notion de **perspectives** ou de **points de vue**. Avant tout, procédons à un petit récapitulatif : nous sommes partis d'un espace-temps vide, dont nous avons vu qu'il n'équivalait pas à un pur néant. Nous avons identifié cet espace-temps vide à une première forme de soi : le **soi témoin**. Puis, par ajout successif d'éléments, nous sommes arrivés à un espace-temps contenant une balle de golf tout d'abord, puis deux balles de golf : une bleue et une rouge.

Imaginons à présent deux moines bouddhistes en pleine méditation : Pierre et Paul (ce sont des moines bouddhistes occidentaux…).

A un temps t1, Pierre et Paul ont réalisé la vacuité. C'est-à-dire qu'ils ont vidé leur esprit de tout phénomène, mental ou physique. Seuls sont présenst le silence et le vide, sous une forme entièrement épurée. Notons bien que dans cet état, il ne saurait y avoir de dualité entre l'observateur (Pierre et Paul) et l'observé (la vacuité). En effet, il ne peut pas y avoir plusieurs manières différentes d'expérimenter le vide. Par conséquent, poser une dualité « observateur/observé » dans le cas de l'expérience du vide n'a guère de sens et relève plus d'une habitude de langage. Il est vrai que dans les cas ordinaires, la dualité « observateur/observé » prend sens, par le jeu de la continuité de souvenirs, de l'orientation volontaire du regard, de l'attitude mentale qui calque une interprétation sur ce qui est observé, etc. Mais cela ne peut pas être le cas à travers l'expérience du vide qui, par définition, se situe en amont du manifesté qu'incarnent, très précisément, les remémorations de souvenirs, l'orientation du regard, la projections de schémas mentaux, etc.

Le vide silencieux dont font l'expérience nos deux moines bouddhistes, et l'expérience qu'ils en ont, ne font donc qu'un. Comment l'expérience du vide pourrait-elle être plus remplie que le vide et être encore « expérience du vide » plutôt que « expérience du vide plus quelque chose » ? Quelle serait la nature de cette différence ? On ne peut répondre à cette question.

A présent, supposons qu'à un temps t2, à travers ce vide, Pierre se représente soudain une balle de golf **bleue** et rien de plus, tandis que Paul se représente une balle de golf **rouge** et rien de plus également.

Puisque Pierre ne se représente pas la balle de golf rouge que Paul se représente, et puisque Paul ne se représente pas la balle de golf bleue que Pierre se représente, il s'ensuit nécessairement que le référentiel à travers lequel elles sont représentées est numériquement différent. Autrement dit, c'est non seulement les deux balles de golf qui sont distinctes (cela va de soi, puisqu'elles sont qualitativement différentes au même moment), mais c'est aussi le vide sur lequel elles se détachent qui n'est pas numériquement le même vide dans les deux cas. C'est, si l'on veut, un vide entourant une balle de golf bleue dans un cas et un vide entourant une balle de golf rouge dans l'autre cas. On a, à présent, deux **perspectives** ou **points de vue** différents qui vont émerger d'une même réalité. On peut dire que le vide se trouve affecté (en creux) dans son identité numérique par la présence d'éléments distincts de lui, bien que qualitativement, il ne change évidemment pas (le vide ne peut être que… vide).

Mais parler d'un changement d'identité numérique est évidemment problématique, puisque le changement présuppose nécessairement la persistance de l'identité numérique. Sans quoi il est une substitution. Or, la substitution implique nécessairement une disparition et une apparition. Et nous avons montré plus haut en quoi ces concepts sont inadéquats en ce qui concerne le vide, l'espace-temps.

La meilleure solution rationnellement acceptable consiste donc à poser que dès t1, le vide n'est pas numériquement un. Il se décompose en autant de vides qualitativement identiques mais numériquement distincts qu'il existe de possibilités différentes de le remplir.

Dans notre expérience de pensée, en t1, le vide réalisé par Pierre, bien que parfaitement indiscernable du vide réalisé par Paul, n'en demeurerait pas moins numériquement distinct, par cela seul qu'il est promis à un destin différent.[287]

Cela nous permet de redéfinir, en le précisant, notre concept de **soi témoin**. Le soi témoin n'est pas simplement l'espace-temps vide. C'est l'espace-temps vide considéré dans sa propension ou sa finalité à être rempli d'une certaine manière. Autrement dit, c'est **un** espace-temps vide, parmi d'autres, qualitativement identiques mais numériquement distincts, constituant chacun un soi témoin différent. C'est la nature et l'arrangement des éléments avec lesquels ils vont s'associer pour former les **perspectives** ou **points de vue** sur les phénomènes du monde réel, qui va distinguer les « soi » témoins les uns des autres,.

Procédons à un bref bilan :

Nous avons mis en évidence l'existence d'un résidu non créable, immuable et indestructible sous-tendant tout fait et tout phénomène ; résidu que nous avons identifié à l'espace-temps et à un **soi témoin** universellement partagé par les êtres.

Nous avons présenté une expérience de pensée suggérant que ce soi témoin, en dépit de sa vacuité, se décline, à travers chaque être, en une multitude de soi témoins numériquement distincts. Nous avons suggéré que lorsque deux êtres, B et C, procèdent d'un troisième, A, cela implique, d'une façon ou d'une autre, une double identité de ce dernier. Cette double identité peut être conçue comme d'emblée présente quoi qu'il arrive.

---

[287] La formulation pourrait laisser croire qu'un événement futur peut modifier un état de fait passé. C'est une interprétation possible, mais elle n'est pas la seule. Cf. notre discussion de la thèse de l'occupation multiple, partie 2 : « Discussion et critique », chapitre 3 : « Des critères aporétiques ».

On peut aussi penser qu'elle apparaît comme effet d'une cause paradoxalement future : la génération des êtres B et C. Ces deux solutions posent des problèmes, déjà abordés lorsque nous avons parlé de la thèse de l'occupation multiple.

Ce qu'il importe de retenir, pour notre présent propos, c'est qu'un concept rudimentaire – mais central – d'identité personnelle peut s'amarrer sur le soi témoin. Et cela est d'une importance capitale pour notre défense de l'analyse des cas fictifs et inhabituels de remise en question de l'identité personnelle. En effet, à partir du moment où le socle même sur lequel fonder le concept de l'identité personnelle est le soi témoin, toute réflexion sur ce concept se doit d'intégrer une réflexion sur le soi témoin. Or, puisque le soi témoin concerne tous les êtres dans toutes les situations possibles et imaginables, une réflexion sur le concept de l'identité personnelle se doit de prendre en compte toutes les situations possibles et imaginables.

# Partie 4 : Conclusion

« Une société d'entités actuelles peut être qualifiée de personne ou d'individu quand la temporalité y prend plus d'importance que la spatialité, autrement dit lorsque la relation avec l'environnement est ressentie comme extérieure, tandis que celle avec le passé est vécue comme un élément constitutif de l'identité […]. La personne est un courant, « une route historique » selon l'expression de Whitehead […], qui commence à la naissance et se termine, peut-être, à la mort. Elle se caractérise par le fait que le lien avec soi est plus fort que le lien avec les autres. »

(André Gounelle, « Le dynamisme créateur de Dieu : Essai sur la théologie du Process »)

Ce travail s'était proposé de décortiquer quelques-uns des problèmes attenants à la notion nébuleuse et pourtant si fondamentale d'identité personnelle.

Nous avons d'abord confronté les deux perspectives les plus opposées sur la question : l'antiréductionnisme, qui prend au sérieux nos intuitions sur le *moi*, et le réductionnisme éliminativiste, qui les sabrent à coup de rasoir d'Occam. Un premier bilan s'est dessiné : si la rationalité empirique échoue à justifier nos intuitions, elle échoue également à les déraciner. Elles restent prégnantes envers et contre tout.

Puis, nous focalisant sur le réductionnisme non-éliminativiste, nous avons exploré les mérites respectifs des critères psychologiques et corporels de l'identité personnelle, dans leurs tentatives de sauver nos précieuses intuitions. Il semble que nous avons pu montrer un léger avantage des critères corporels sur les critères psychologiques dans cette tâche.

Enfin, nous avons suggéré une possibilité de « grande réconciliation » de nos intuitions et des faits, venant d'une prise en compte d'un grand oublié des discussions sur le *moi* : son environnement. En inscrivant chaque individu dans un réseau relationnel[288] qui le rend unique, de par la place qu'il y occupe ; qui confère un sens transcendant à son existence, au regard de sa relation au réseau ; et qui lui garantit une identité à travers le temps, pour autant que les déformations de ce réseau n'en modifie pas la topologie ; le contexte apparaît être la version moderne de l'âme.

[288] Physique et social.

Notre discussion s'est achevée par une présentation critique d'arguments sceptiques importants opposés à l'usage des expériences de pensée dans le cadre des réflexions sur le *moi*. Nous suggérons que ce scepticisme provient d'une sous-évaluation de l'enjeu métaphysique de ces réflexions sur le *moi*, qui dépassent le cadre d'une simple analyse conceptuelle, pour toucher à la nature même de la réalité.

Récapitulons à présent les grands problèmes que nous avons rencontrés, par rapport à l'identité personnelle, et voyons quelle réponse nous pouvons leur apporter, à l'éclairage de notre discussion (nous utiliserons quelquefois la première personne du singulier par commodité). Notre thèse a essentiellement porté sur le problème de l'identité à travers le temps, toutefois, nous pensons qu'avoir abordé le problème sous cette perspective nous a permis de dégager malgré tout des pistes intéressantes au regard d'autres perspectives. Voici les principales grandes questions que nous nous sommes posées au début de ce travail :

**1°) IDENTITE SPECIFIQUE ET NUMERIQUE : Qu'est-ce qui fait de moi *une* personne ? *cette* personne ?**

Critère physique : En premier lieu, nous sommes *une* unité physique dynamique, structurelle, fonctionnelle, richement autorégulée, téléonomique, en bref : un corps vivant[289] – *ce* corps – doté d'*un ensemble* de caractéristiques objectives intrinsèques (masse, volume, morphologie, physiologie, etc.) – *ces* caractéristiques – formé d'une certaine fraction de matière – *cette* fraction de matière – localisée en un certain espace – *cet* espace –

[289] Par corps vivant, il faut entendre ici une notion très large, incluant certes les corps animaux, étudiés par la biologie, mais aussi, pourquoi pas, des corps artificiels de robots qui seraient suffisamment perfectionnés pour posséder une sensibilité.

inséré dans un certain contexte physique – *ce* contexte – ce qui fait que nous nous distinguons d'*une autre* personne existant au même moment et qui serait intrinsèquement indiscernable, mais qui, nécessairement, serait formée d'*une autre* fraction de matière localisée en *un autre* espace, insérée dans *un autre* contexte physique. Si nous sommes des êtres non isolés, et dotés d'une conscience de soi en première personne, alors un tel critère est une condition nécessaire à notre identité, mais pas un critère suffisant. Si nous sommes des êtres isolés[290] et dépourvus de conscience de soi en première personne[291], alors un tel critère est une condition nécessaire et suffisante à notre identité.

Critère social : En second lieu, outre notre insertion dans un contexte physique, il faut tenir compte également, **si elle existe**, de notre insertion dans un contexte social (*ce* contexte social). Cette situation particulière porte atteinte, cognitivement, épistémologiquement et ontologiquement, à la singularité de notre ego (qui va de soi lorsqu'on est seul), en nous mettant face à un autre nous-même qui n'est pas *moi*. D'un autre côté, comme nous l'avons vu, autrui participe – cognitivement sans doute, épistémologiquement c'est probable, et ontologiquement, peut-être – à la constitution de notre *moi*, par sa reconnaissance de notre identité, laquelle passe avant tout par la prise en

---

[290] Il nous apparaît important de préciser que si notre insertion dans un contexte social est constitutif de notre identité, cela n'est vrai que si un tel contexte social existe. Mais son absence n'implique pas nécessairement que l'on n'est pas une personne. Ainsi, le dernier survivant après l'apocalypse, bien qu'isolé, serait sans doute, encore, une personne.

[291] Il nous apparaît important de préciser que si la conscience de soi en première personne est constitutive de notre identité personnelle, cela n'est vrai que dans la mesure où une telle conscience de soi existe. Mais son absence n'implique pas nécessairement que l'on n'est pas une « personne », en un sens très large du terme. On peut, par exemple, soutenir qu'un animal, dépourvu de conscience de soi, pourrait être quand même une « personne », au sens très large du terme, en ce sens qu'il demeurerait un foyer de sensibilité (à son corps et à son environnement), autocentré. De même, le fait de dormir, et donc de perdre provisoirement conscience de soi, n'implique pas que l'on cesse d'exister en tant que personne.

considération de notre corps, et après seulement, de notre psyché (bien que tout un chacun se reconnaît comme le même, d'abord par sa psyché, et ensuite seulement par son corps). L'insertion dans un contexte social est une condition nécessaire à notre identité, mais encore une fois, ce n'est pas une condition suffisante, sauf si nous sommes des êtres dépourvus de conscience de soi en première personne.

En troisième lieu, nous sommes, **aussi**, **s'ils existent,** *un certain nombre* de souvenirs – *ces* souvenirs – de traits psychologiques – *ces* traits psychologiques – une certaine personnalité – *cette* personnalité – plus précisément : *une* conscience de soi en première personne, qui est capable d'effectuer des choix selon des préférences – *cette* conscience de soi, etc. Cette conscience de soi est insérée dans *un certain* contexte culturel voire linguistique (*ce* contexte culturel et linguistique) qui participe de son identité. Cela est aussi une condition nécessaire, mais pas suffisante.

Il semble que plusieurs personnes peuvent cohabiter dans le même individu (ex. : jumeaux siamois, patients split brain, syndrome de personnalité multiple, etc.). En revanche, une même personne ne peut pas être incarné par plusieurs individus (successivement ou en même temps).

**Synthèse :**

**Etre *une* personne, c'est être – au moins une partie d' – *un* individu matériel vivant, spatialement situé, voire socialement situé, voire : culturellement situé et capable de manifester *une* conscience de soi en première personne et d'effectuer des choix selon *des* préférences.**

**Etre *cette* personne, c'est être *cet* individu matériel vivant – ou *cette* partie de celui-ci – situé dans *cet* espace, voire *ce* contexte social, voire : *ce* contexte culturel et capable de manifester *cette* conscience de soi en première personne** (du moins, un noyau plus ou moins stable de pensées égologiques présentes dans cette conscience de soi) **et d'effectuer des choix selon *ces* préférences.**

### 2°) IDENTITE DIACHRONIQUE : Jusqu'à quel point puis-je changer ?

Bizarrement, bien que la fraction de matière qui nous constitue et l'espace que l'on occupe sont essentiels pour notre identité à un moment donné, il semble bien que seule une continuité de matière et de localisation spatiale soit requise pour notre identité à travers le temps.[292] Ainsi, il va de soi que notre position dans l'espace change très souvent. Il faut savoir, en outre, qu'en une dizaine d'années, notre corps est entièrement renouvelé. Il ne contient plus rien de la matière qui le constituait dix ans plus tôt. Les expériences de pensée de téléportation jettent un doute sur la continuité matérielle elle-même, et semblent privilégier un critère psychologique de l'identité à travers le temps. Il va de soi que nous aurions du mal à reconnaître comme la même personne, quelqu'un qui aurait subi un traumatisme mental ou un changement de personnalité sévère (« Ce n'est plus le même homme ! », dirons-nous souvent). Mais en revanche, il est incontestable que deux individus numériquement distincts, mais qui auraient exactement la même personnalité, les mêmes souvenirs, les mêmes traits psychologiques, seraient jugés être deux personnes différentes,

[292] Une continuité de contexte social et culturel n'est pas indispensable. Le dernier survivant après l'apocalypse serait la même personne qu'avant l'apocalypse. En revanche, une continuité spatiale et temporelle, quoique nécessaire, n'est pas suffisante à la préservation de l'identité. Ainsi, bien que Hyde soit en continuité spatiale et matérielle avec Jekyll, nous estimons que ce sont deux personnes numériquement distinctes.

et non la même personne en deux exemplaires ou à deux endroits à la fois. Par contre, on peut certainement concevoir que deux personnes numériquement distinctes existent – simultanément ou successivement – dans le même corps (ex. : jumeaux siamois, patients split brain, syndrome de personnalité multiple, amnésies profondes, etc.). En outre, nous concevons mieux (au moins scientifiquement) une personne réduite à un corps vivant (fut-ce provisoirement, par exemple lors du sommeil profond), qu'une personne sans corps qui serait hypothétiquement réduite à un esprit… Une personne ne peut pas juste exister sous un mode abstrait, désincarné. Si tel était le cas, il se pourrait qu'elle puisse informer plusieurs corps numériquement distincts (en même temps ou successivement). Ainsi, il faudrait dire qu'une même personne existe à différents endroits à la fois, à travers plusieurs corps. Nous avons montré comment une telle option était problématique.

Bien sûr, ce problème se pose aussi, dans une moindre mesure, si une personne est d'abord un corps matériel situé dans un certain espace, comme l'illustre le cas des split brain, ou celui de la double transplantation de cerveau. Nous avons esquissé trois solutions en accord avec nos intuitions antiréductionnistes concernant l'identité personnelle.

La première solution – même si nous l'avons évoquée en troisième dans notre thèse, nous la plaçons ici en premier, car ce n'est pas elle que nous retiendrons – fonctionne bien dans un cas précis : le cas où une action autonome est impliquée dans un processus de fission, sans en être la cause. Il s'agit de la solution que nous avons tirée de l'hypothèse de David Banach. Selon cette solution, la personne s'identifie – cognitivement, mais aussi ontologiquement – avec l'individu néoformé qui actualise le mieux son intention de départ.

La deuxième solution fait adhérer l'ontologie à l'épistémologie, voire à la cognition, notamment sociale. C'est une solution assez paradoxale, il faut bien l'avouer, qui repose sur un fort présupposé externaliste. Elle consiste à dire que le maintien de l'identité d'une personne au cours du temps, dépend étroitement de ce qu'une majorité de personnes aux facultés cognitives normales (sans chercher ici à définir ce terme…) en penseront ou en penseraient. C'est-à-dire, si l'on considère deux personnes X et Y vivant à des moments différents : Y est la même personne que X si et seulement si une majorité de personnes aux facultés cognitives normales (Y y compris) ayant connues X s'accordent ou s'accorderaient à reconnaître Y comme étant X. Nous ne retiendrons pas non plus cette solution, car il ne paraît guère vraisemblable qu'un tel critère soit une condition nécessaire à l'identité personnelle, même s'il n'est pas invraisemblable qu'il en soit une condition suffisante. Ainsi, supposons qu'une majorité de gens s'accordent à identifier Y avec X. Il s'ensuit que X est **réellement** Y. Mais imaginons que cette majorité de gens s'accordant à identifier Y avec X ne se joue qu'à une ou deux personnes, et que ces dernières, finalement, décèdent ou changent d'avis, si bien qu'à présent, une majorité de personnes n'identifient pas Y avec X. Il faudrait en conclure que maintenant, X n'est **réellement** plus Y. Mais si X est numériquement identique à Y, dans le monde **actuel**, alors il est nécessairement identique à Y dans ce monde. De même que Phosphorus ne peut pas cesser d'être Hespérus.

La troisième solution est la thèse de l'occupation multiple. Là où l'on croit voir une seule personne, il y a en fait plusieurs personnes numériquement distinctes, quoique qualitativement identiques en temps ordinaire. Dans le cas, par exemple, de la double transplantation de cerveau, elles sont amenées à vivre séparément et à se différencier qualitativement. Cette solution fonctionne relativement bien dans les deux cas suivants, qui seront autant de versions de cette solution : Le cas où c'est un processus rigoureusement

déterministe qui donne lieu aux fissions d'un seul individu en deux (ou plusieurs) individus corporellement en continuité avec le premier. Le cas où l'on admet qu'un individu puisse être plusieurs personnes différentes en même temps, même si cet individu ne subit aucune fission durant son existence.

Dans le premier cas, nous pouvons admettre que la plupart des individus actuels sont une seule personne à la fois, conformément à nos intuitions communes. Seuls les individus qui subiront un jour une fission sont plusieurs personnes à la fois (autant que d'individus qui seront engendrés). Puisque le processus serait rigoureusement déterministe, cette rétro-causalité pourrait n'être qu'apparente : par exemple, l'enchaînement d'événements qui va conduire à une fission, ainsi que la multiple identité personnelle de l'individu qui va fissionner pourraient avoir une cause commune.

Dans le second cas, on est obligé d'admettre que tout individu est, dès le commencement de sa carrière, plusieurs personnes (autant que d'individus qui *pourraient être* générés par fission de cet individu). Malheureusement, ce nombre semble, de prime abord, virtuellement infini, puisque chaque individu généré par fission de l'individu de départ, pourrait très bien, lui aussi, fissionner…

On peut donc préférer la première version de la solution, mais elle ne marche bien que si on postule que toutes les fissions sont dues à un processus strictement déterministe. Cependant, rien ne peut le garantir.

En définitive, c'est sans doute la seconde version de la solution la plus crédible. En effet, la possibilité d'avoir une infinité d'individus, obtenus par fissions successives d'un

individu, et qui soient également en continuité corporelle avec ce dernier, du moins selon un mode adéquat, est peu plausible.

Voici pourquoi : le processus de fission implique que l'individu de départ subisse un certain nombre de contraintes structurelles telles qu'il n'est pas dit qu'il soit légitime de parler de continuité corporelle, du moins, au-delà de quelques fissions. Dans le cas, par exemple, où un individu se diviserait comme une amibe en deux individus, passant par une série d'étapes de « pseudo frères siamois », partiellement mais incomplètement dédoublés, on peut supposer que cette entité intermédiaire serait bien trop structurellement déformée par rapport à l'individu de départ, pour que nous puissions parler légitimement de continuité corporelle, et donc de préservation de l'identité personnelle. Et à supposer qu'il soit légitime de parler de continuité corporelle, en revanche, il n'est pas du tout clair que le type de continuité corporelle impliqué permette de préserver l'identité numérique de la personne (contrairement au renouvellement naturel des cellules du corps au cours de la vie). Ainsi, admettons qu'un individu se divise (progressivement, si l'on veut) en deux moitiés d'individus, puis, seulement, que dans un second temps, ces deux moitiés reconstituent chacune la partie manquante, de manière à former deux individus parfaitement indiscernables. Environ une moitié seulement des tissus de chacun des deux individus résultants sera constituée des tissus originaux de l'individu de départ. Les autres tissus seront, au mieux, engendrés à partir de ces moitiés, et, même si ce processus de complétion est très progressif[293], il pourrait être tout simplement incompatible avec la préservation de l'identité personnelle. Après tout, la formation de l'embryon dans l'utérus de sa mère est aussi très progressive, et les éléments utilisés pour son développement sont, à la très notable exception du spermatozoïde ayant fécondé l'ovule, puisés dans le corps de

[293] Mais deux individus incomplets seraient-ils viables suffisamment longtemps ?

la mère. Pourtant, qui douterait que l'enfant est numériquement distinct de sa mère ? A la fois comme personne, mais aussi, tout d'abord, comme individu. Par conséquent, nous avons là un exemple flagrant de continuité corporelle qui n'entraîne pas une préservation de l'identité numérique de la personne. Aussi, il est possible que lors d'une hypothétique division d'un individu pour en donner deux qualitativement identiques, seule la moitié de chaque individu résultant entretienne, avec l'individu de départ, le type de continuité corporelle requis pour la préservation de l'identité personnelle numérique. L'autre moitié étant quelqu'un d'autre de nouveau.

Pour y voir plus clair, on peut utilement distinguer deux types de **continuité corporelle** :

- Une **continuité régénératrice**, compatible avec le maintien de l'identité personnelle : qui concerne le renouvellement et la réparation du corps, ainsi que son développement, sa croissance, mais aussi sa vieillesse, sa détérioration, et qui affecte pas ou peu l'organisation anatomique et physiologique ;

- Une **continuité formatrice**, incompatible – semble-t-il de prime abord – avec le maintien de l'identité personnelle : qui concerne la création, de novo, de structures préalablement inexistantes, ou ayant été détruites, à partir de ce qui existe déjà et/ou encore ; qui concerne aussi des destructions massives de structures ; et qui affecte, en tout cas, beaucoup l'organisation anatomique et physiologique. (ex. : la queue du lézard qui se reforme après avoir été sectionnée, l'enfant qui se forme dans l'utérus de

sa mère, l'étoile de mer qui, coupée en deux, reforme deux étoiles de mer complètes, etc.)

Bien sûr, la différence entre ces deux types de continuité n'est pas toujours claire. La croissance et le développement impliquent aussi, parfois, et surtout au début, l'apparition de structures nouvelles. Les dents adultes qui viennent remplacer mes dents de laits ne sont-elles pas véritablement à moi, mais à quelqu'un d'autre ? De plus, ajouter au corps une nouvelle cellule, ou remplacer une cellule détruite par une nouvelle cellule, n'est-ce pas, déjà, un cas de continuité formatrice ? Ainsi, il s'agirait plus, avec ces deux types de continuité corporelle, d'envisager les deux pôles d'un même axe, que d'envisager deux types de continuité corporelle radicalement distincts.

On peut, semble-t-il, fixer un second critère qui permette de distinguer, cette fois-ci de manière plus radicale, entre deux types de continuité formatrice :

- Une **continuité formatrice sub-personnalisante**, compatible avec la préservation de l'identité personnelle, car elle concerne la création/destruction de structures dont la taille, l'organisation, le fonctionnement, le degré de développement, etc. ne les rendent pas propres à servir de base à l'existence d'une personne (ex. : dents, neurones isolés ou petits réseaux de neurones, os, muscles, vaisseaux sanguins, etc.) ;

- Une **continuité formatrice personnalisante**, incompatible avec la préservation de l'identité personnelle, car elle concerne la

création/destruction de structures suffisantes pour faire exister une personne véritable. (ex. : cerveaux, hémisphères cérébraux, larges pans de tissus neuronaux, etc.)

Avec cette distinction en tête, reprenons notre cas de duplication par fission et reconstitution de la partie manquante. Admettons que ce processus fonctionne correctement, on a donc, au départ, un individu. Cet individu est formé de deux moitiés symétriques qui sont, chacune, une personne. Les moitiés se séparent. Les deux personnes constituant l'individu de départ se trouvent donc séparées. Puis, chaque moitié d'individu reconstitue la moitié manquante, de manière à former deux nouveaux individus. Chaque nouvelle moitié réunie les conditions suffisantes pour être le support à l'existence d'une personne. Chaque nouvelle moitié est donc en **continuité formatrice personnalisante** avec la moitié de l'individu de départ qui lui a donné naissance. Par conséquent, chaque nouvelle moitié sera le support de l'existence d'une personne, et cette personne sera entièrement nouvelle (du point de vue de l'identité numérique).

Si jamais le processus s'était avéré défectueux, et que les moitiés nouvellement générées n'avaient pas eu d'hémisphère cérébral, par exemple, alors chacune aurait été en continuité formatrice sub-personnalisante avec la moitié de l'individu de départ lui ayant donné naissance. Par conséquent, ces moitiés néoformées n'auraient pas été le support de nouvelles personnes, elles n'auraient été que des prolongements corporels des personnes, déjà existantes, et correspondant aux moitiés originales de l'individu de départ.

**Synthèse : Pour B, être la *même* personne que A, c'est être – au moins une partie d' – un individu matériel, spatialement situé, voire socialement situé, voire :**

**culturellement situé et capable de manifester une conscience de soi en première personne et d'effectuer des choix selon des préférences, tel que les deux conditions suivantes soient réalisées :**

- **B est en *continuité corporelle régénératrice* ou *formatrice sub-personnalisante* avec l'individu matériel vivant – ou la partie de celui-ci – qui est le support physique de A ;**

- **La conscience de soi de B contient un certain noyau plus ou moins stable de pensées égologiques qui sont aussi présentes chez A.**

**3°) UNITE DE LA CONSCIENCE ET DE L'EXPERIENCE : Comment concevoir le passage de une conscience[294] à plusieurs, de plusieurs consciences à une seule ? Y a-t-il un moment précis où il s'effectue ? Quelle est la nature d'une entité consciente dont les bases physiques sont partiellement mais pas entièrement dédoublées ? Comment s'y organisent les contenus expérientiels ?**

Le substantialisme et l'internalisme montrent de sérieuses limites à interpréter l'unité de la conscience et de l'expérience, dès lors où l'on s'intéresse à des cas qui semblent impliquer un dédoublement partiel de celles-ci.

Nous avons montré qu'une ontologie du processus, conjointe à la thèse de l'externalisme actif, permettait une interprétation satisfaisante. En effet, elles rendent

[294] Pour ne pas alourdir le texte, nous confondons ici ce qui relève du champ d'expérience et ce qui relève de la conscience proprement dite sous le terme « conscience ».

possible le partage, entre deux centres de conscience numériquement distincts, de réalités conscientes et phénoménales numériquement identiques (du moins quant à leur soubassement physique). En outre, l'externalisme permet de comprendre qu'une unification ou dissociation de contenus conscients puisse se réaliser sur la base d'une unification ou dissociation de processus physiques *externes*, et pas seulement *internes*. Ainsi, le mystère des patients split brain, tantôt semblant être une seule conscience[295], tantôt semblant être deux consciences[296], est résolu.

Le passage à une ontologie du processus et à l'externalisme actif a une conséquence plus générale au sujet de l'identité personnelle. Nous avons dit qu'être une personne, c'était être une certaine fraction de matière (entre autres), qu'être cette personne, c'était être cette fraction de matière (entre autres), et que l'identité diachronique impliquait (entre autres), une continuité matérielle. Nous n'avons pas, par là, voulu signifier un engagement ontologique envers le substantialisme, mais simplement insister sur l'importance de la dimension ontologique de la corporéité dans la définition de l'identité personnelle. Avec l'éclairage d'une ontologie du processus, nous pouvons reformuler les choses ainsi : être une personne, c'est être un certain processus (entre autres), être cette personne, c'est être ce processus, et l'identité diachronique implique (entre autres), une continuité processionnelle.

**4°) IDENTITE SYNCHRONIQUE : Jusqu'à quel point aurais-je pu être différent de ce que je suis ? Aurais-je – subjectivement – pu être une autre personne – objectivement ?**

---

295 Dans la vie de tous les jours, réalisant la connexion de leurs hémisphères par le biais de processus physiques externes impliquant leur corps et leur environnement.

296 Lorsque des situations expérimentales particulières viennent interrompre les processus physiques externes d'intégration évoqués dans la note 295.

En introduction, nous avons évoqué le problème de l'identité personnelle à travers les mondes possibles, et avons précisé que nous ne le développerions pas. Dans notre travail, nous n'y sommes, effectivement, pas revenu directement. Toutefois, nous l'avons malgré tout abordé incidemment, lorsque nous nous sommes progressivement plongé dans des considérations de plus en plus métaphysiques, à travers la partie 3, « Identité personnelle et expériences de pensée », et notamment lorsque nous avons commencé à réfléchir à la notion de « soi » témoin. Toutes ces réflexions sont, en effet, très liées.

Je suis, actuellement, *ce* corps, formé de *cette* fraction de matière, et situé dans *cet* espace, etc. Aurais-je pu avoir *un autre* corps, formé d'*une autre* fraction de matière, et situé dans *un autre* espace, etc. ? On croit pouvoir l'imaginer. On croit même pouvoir imaginer avoir été quelqu'un d'autre parmi les milliards (si on se limite à la planète Terre) de personnes existantes.

Etant donné que la matière qui constitue notre corps est appelée à se recycler, on a du mal à imaginer que la matière nous constituant à l'origine soit une composante essentielle de notre identité (après une dizaine d'années de vie, nous n'en avons plus la moindre trace).

Nous changeons continuellement de localisation dans l'espace. Au lieu d'être allé à Lille aujourd'hui, vous auriez pu aller à Marseille. Pourquoi le fait d'être né à Lille plutôt qu'à Marseille devrait donc être une composante essentielle de notre identité ? Si, lorsque vous êtes né, votre mère avait été en déplacement dans un autre lieu que celui dans lequel elle a réellement accouché, cet enfant n'aurait-il pas été vous pour autant ? Cela semble peu crédible.

Il semble que l'ovule et le spermatozoïde qui nous ont donné naissance auraient pu être mis en congélation, et seulement unis et implantés dans un utérus au bout de 10 ans, et que nous serions alors nés 10 ans plus tard. Le moment de notre conception ne semble donc pas non plus être essentiel.

C'est un fait : au début de notre vie, notre conscience de soi en première personne était très rudimentaire. Mais, avec les années, elle s'est enrichie. Et c'est pourquoi nous pouvons avoir du mal à croire que nos pensées égologiques sont essentielles à notre identité, alors qu'une vie différente aurait pu nous en faire acquérir d'autres.

Si, dans notre monde actuel, il parait indéniable, à la lumière des points 1°), 2°) et 3°), que nous sommes une certaine personne relativement bien définie, avec des critères d'identité diachronique relativement bien définis, il semble bien que la poursuite du raisonnement précédent nous conduirait facilement à tout remettre en question. Nous conclurions sur l'idée d'une identité personnelle qui ne serait qu'un leurre – à la manière des réductionnistes éliminativistes – ou bien serait décidemment irréductible (au moins ontologiquement) à aucun fait empirique – à la manière des antiréductionnistes. Car il semble bien que nous aurions pu être n'importe qui d'autre, de complètement différent de qui nous sommes. Si nous recherchons le point commun à tout ce que l'on aurait pu être, il semble que l'on retombe finalement sur ce « soi » minimal qu'est le « soi » témoin, analysé au cours de notre partie 3, « Identité personnelle et expériences de pensée ».

Nous avons suggéré, toutefois, que ces impressions d'avoir été quelqu'un d'autre, ou d'avoir été radicalement différent, pouvaient bien être en grande partie trompeuses. Si cela

est correct, alors il y a davantage d'aspects nécessaires, dans ce que nous sommes, que ce que ce genre d'exercice de l'imagination peut nous faire croire.

Nous avons aussi suggéré une manière de donner un sens à ces impressions, s'appuyant sur l'expérience de pensée de la désincarnation progressive. Ainsi, il serait métaphysiquement possible d'avoir été quelqu'un d'autre de radicalement différent, si nous étions, essentiellement, une sorte d'« esprit ». Pas, au sens du critère psychologique, un ensemble de souvenirs, de traits psychologiques, de pensées conscientes en première personne. Mais plutôt en un sens quasi-religieux : un être bien vivant d'une certaine façon, bien individualisé, mais invisible, intangible et dépourvu de corps, dont l'existence consisterait à errer dans le monde et parmi les vivants, à observer ce qui s'y passe sans pouvoir y agir. Un peu comme l'ange Damiel qui, dans les *Ailes du désir*[297], parvient à capter toutes les pensées des mortels, mais demeure incapable d'infléchir les sentiments dépressifs d'un suicidaire qui finit par se jeter dans le vide. Cet « esprit », bien entendu, n'est pas le « soi » témoin, puisqu'il dispose d'une conscience de soi en première personne, de préférences, de motivations, etc. Mais la prise en considération de cet « esprit » – qu'il existe ou non – permet de faciliter la compréhension du « soi » témoin, par notre variante de l'expérience de pensée de la désincarnation progressive, qui montre comment l'on peut être pleinement quelqu'un d'autre, et même plusieurs personnes successivement, sans réellement cesser d'être soi-même. Le « soi-même » en question, en ce qui concerne le « soi » témoin, est bien sûr réduit à sa plus simple expression.

Ces perspectives hypothétiques semblent nous emmener bien loin de ce que nous avons pu dire précédemment. Du reste, le simple fait qu'un scénario soit conceptuellement,

---

297 HANDKE Peter et WENDERS Wim. *Les Ailes du désir.*

voire métaphysiquement, possible n'implique pas, on l'a vu, qu'il soit métaphysiquement vrai. Comme nous l'avons déjà dit, on peut imaginer qu'en bougeant ses deux bras on va réussir à s'envoler. C'est sans doute conceptuellement et métaphysiquement possible, mais cela demeure faux – et physiquement impossible. Par contre, dans le cas du « soi » témoin, les expériences de pensée que nous avons menées n'ont pas eu pour but d'exhiber une possibilité conceptuelle, dont on revendiquerait la réalité – et non simplement la possibilité – métaphysique, mais une nécessité conceptuelle, dont nous avons revendiqué la nécessité métaphysique. Cette nécessité conceptuelle du « soi » témoin équivaut à l'impossibilité conceptuelle du néant subjectif (c'est-à-dire, une impossibilité d'imaginer sa propre inexistence en première personne). Nous avons montré, dans notre partie 3 : « Identité personnelle et expériences de pensée », que la nécessité conceptuelle n'entraînait pas logiquement la nécessité métaphysique (alors que la nécessité métaphysique entraîne la nécessité conceptuelle). Toutefois, nous pensons que cette idée générale doit être nuancée au cas par cas. En l'occurrence, nous estimons que la nécessité conceptuelle du « soi » témoin implique logiquement une nécessité métaphysique, ou, autrement dit, que l'impossibilité conceptuelle du néant subjectif implique une impossibilité métaphysique.[298]

ꕥ

[298] Dans les annexes, nous développons un certain nombre d'arguments pour étayer cette dernière idée.

# Annexe : Identité personnelle et survie

« La mort n'est pas un événement de la vie. On ne vit pas la mort.

Si on entend par éternité non la durée infinie mais l'intemporalité, alors il a la vie éternelle celui qui vit dans le présent.

Notre vie n'a pas de fin, comme notre champ de vision est sans frontière. »

(Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, § 6.4311)

# I. Introduction

Comme nous l'avons déjà esquissé dans le corps de cette thèse, une réponse à la question de l'identité personnelle qui satisfasse tous les cas possibles peut se justifier d'un point de vue métaphysique. A partir du moment où l'on considère qu'une description n'a pas seulement un rôle instrumental mais qu'elle vise aussi à refléter une réalité sous-jacente à l'expérience sensible, alors rechercher la vérité, qui est une motivation tout à fait légitime en soi, exige qu'une description permette de rendre compte, non seulement de tous les cas réels/habituels mais aussi de tous les cas qui pourraient potentiellement se présenter. En science, si une hypothèse n'explique pas un fait ou est contradictoire avec l'existence de ce fait, alors c'est que cette hypothèse est fausse, ou plus précisément incomplète, et qu'une autre hypothèse plus générale devra être trouvée qui rende compte, non seulement des mêmes faits que l'hypothèse précédente, mais aussi des faits que l'hypothèse précédente ne permet pas d'expliquer. En philosophie, lorsqu'on ne peut pas à proprement parler effectuer des vérifications empirico-expérimentales des hypothèses, on peut au moins tenter une confrontation de celles-ci avec : l'expérience subjective, d'autres hypothèses que l'on tient pour vraies parce qu'elles ont été scientifiquement vérifiées ou parce qu'elles s'imposent intuitivement (l'écoulement du temps, l'existence de l'espace et du monde extérieur, l'existence d'autrui , etc.) mais qui pourraient être fausses/incomplètes, d'autres hypothèses/croyances invérifiables scientifiquement et qui ne s'imposent pas à notre intuition (temps linéaire ou temps cyclique, existence d'univers parallèles, espace-temps divisible ou non à l'infini, etc.), ce qui ne permet pas vraiment de statuer sur la valeur de vérité de ces hypothèses, mais permet au moins quelques conditionnels du style : si

l'hypothèse A est vraie alors l'hypothèse B est vraie également mais l'hypothèse C est fausse, etc.

En particulier, apporter une réponse à la question de l'identité personnelle est intéressant – nous l'avons également évoqué dans le corps de cette thèse – pour répondre à une autre question dont on peut difficilement nier le caractère ontologique et l'intérêt : la question de l'après-vie.

La question de l'identité personnelle est en effet très liée à la question de l'après-vie, de l'immortalité de l'âme, etc. C'est d'ailleurs, historiquement, dans ce contexte que le problème de l'identité personnelle fût posé pour la première fois. Le type de réponse que l'on pourrait apporter à la question de l'identité personnelle aura des répercussions quant à la possibilité qu'il existe ou non une vie (éternelle) après la mort. En effet, supposons que la meilleure manière que nous ayons de répondre à la question de l'identité personnelle consiste à émettre l'hypothèse que la personne est essentiellement une entité non-physique (une âme par exemple), alors cela s'accorderait bien avec l'idée selon laquelle on continue d'exister après la mort physique. Ce ne serait pas une certitude absolue mais cela deviendrait très plausible car une entité non-physique n'a pas de raison d'être affectée par un processus de dégradation physique. A l'inverse, si la meilleure manière que nous ayons de résoudre le problème de l'identité personnelle consiste à identifier la personne avec un corps matériel, alors cela sera difficilement compatible avec une survie personnelle, mais pas forcément avec une survie trans-personnelle, une continuité dans la subjectivité non liée avec une continuité de la personne[299].

[299] Cf. CLARK, Thomas W. "Death, Nothingness, and Subjectivity"

Apporter un fort élément de réponse à la question de l'après-vie est selon nous une des motivations les plus fortes à trouver un critère satisfaisant de l'identité personnelle. Cela ne semble pourtant pas aller de soi pour tout le monde. Selon certains, savoir s'il y a ou non une vie après la mort serait sans importance car « on verra bien au moment venu ». Pour d'autres, relativement bien convaincus qu'il n'y a rien à espérer de ce côté-là, le néant est ce qu'il peut nous arriver de mieux car seule une vie limitée dans le temps lui donnerait tout son sens (car il faudrait cueillir le présent et ne pas remettre à plus tard) et fonderait la morale (il n'y aurait pas de mal possible si rien ne pouvait marquer un être de façon irrémédiable). Il semble au contraire que la perspective du néant rend la vie désespérante, absurde et angoissante.

Désespérante, car tout projet devient vain, marqué du sceau de l'éphémère. A quoi bon se donner de la peine pour se cultiver, s'améliorer moralement, etc. si tous ces efforts sont en pure perte sur le long terme ? Autant profiter hédonistiquement de l'instant présent sans se casser la tête.

Absurde, car notre existence devient un fait brut, sans raison profonde. Nous sommes là mais nous aurions très bien pu ne pas être là et ça n'aurait pas changé grand'chose. De plus, quand nous ne serons plus, ce sera exactement comme si nous n'avions jamais été, donc pourquoi vivre ? On pourrait se dire : nous sommes là, autant en profiter. Oui, mais celui qui se suiciderait ne ferait pas un meilleur ou un moins bon choix. Une fois réduit à néant, il ne se rappellerait évidemment pas avoir vécu, et il n'éprouverait aucun remord. A la limite, des considérations morales, du moment qu'on est vivant, pourraient faire préférer la vie à la mort. On ne veut pas forcément faire souffrir ses proches. Mais une fois mort, de telles préoccupations morales ne nous préoccuperaient

plus. De plus, on pourrait, au moins en théorie, imaginer de tuer l'ensemble de l'humanité en même temps, sans souffrance et instantanément. En ce cas, la souffrance des proches ne serait plus un argument contre le suicide puisqu'il n'y aurait personne pour pleurer la disparition de toute l'humanité ! Bien sûr, dire une telle chose peut rebuter instinctivement, mais une réflexion éthique saine montrerait qu'il n'y aurait aucun mal à commettre un tel génocide indolore et instantané si on suppose que la mort débouche sur le néant. Certains objectent que l'on aurait en fait au moins une bonne raison de vivre : on vivrait pour la génération suivante. Mais elle-même, pour qui vivrait-elle ? En toute logique, pour la génération suivante également, et ainsi de suite jusqu'à la dernière génération d'êtres humains qui elle ne vivrait pour personne puisqu'il n'y aurait plus personne après. Par récursivité, on peut donc dire que vivre pour la génération suivante, n'est pas une bonne raison de vivre.

Enfin, l'existence serait angoissante. Le déclin et la disparition irrémédiable des gens que l'on aime est quelque chose, peut-être même la seule, à mériter véritablement le qualificatif de tragique. De plus, l'inéluctabilité de la mort nous pousse à vouloir vivre au maximum et on peut avoir l'impression de ne jamais véritablement réussir à goûter pleinement à l'instant présent, qui nous échappe sans cesse pour rejoindre le passé. Le bonheur présent entraîne la crainte d'un malheur futur et le bonheur passé suscite le regret. Mais surtout, le néant de la conscience est quelque chose qui nous échappe, qu'on ne peut définir adéquatement. On ne sait pas quel effet cela fait de ne plus être, et cette pensée peut nous angoisser.

Une objection qui semble évidente à ce point est que dans les faits, bien peu de gens connaissent une véritable angoisse de vivre. Bien peu, parmi les croyants en l'immortalité

personnelle, oui, on peut bien le croire. En est-il de même parmi les athées et sceptiques ? On peut en douter, du moins dans un certain nombre de cas. D'ailleurs, parmi les athées et les sceptiques qui ne connaissent pas d'angoisse de vivre, combien ont poussé leur réflexion suffisamment loin pour se rendre compte de ce qu'implique une existence limitée dans le temps ? Enfin, il faut aussi considérer le fait que face à une même situation potentiellement anxiogène, tout le monde ne réagit pas forcément de la même façon, et ce à cause de facteurs individuels (meilleure maîtrise de ses émotions, sécrétion anormalement basse d'adrénaline, etc.) ou contextuels (rythme de vie intense ou au contraire serein, vie sociale et sentimentale riche et épanouissante, etc.) mais il suffit que la perspective du néant soit angoissante pour au moins une personne et que cette personne soit réfractaire à tout ce qui ne serait pas une démonstration rationnelle d'une vie après la mort pour que cela vaille la peine de tenter d'effectuer cette démonstration.

Loin d'être un bien, la perspective du néant gâche donc notre vie. Démontrer que ce néant est impossible et mieux, qu'une survie éternelle et personnelle nous attend après la mort est donc la meilleure chose qui puisse nous arriver. Bien sûr, la tâche est peut-être vaine, car il n'y a peut-être rien à démontrer, soit parce que la proposition est indémontrable, soit parce qu'elle est fausse, mais, si elle est vraie, alors c'est probablement la chose la plus utile qu'il conviendrait de savoir et, s'il y a la plus petite chance pour qu'elle soit vraie et démontrable, alors il faut tout mettre en œuvre pour la démontrer. S'il fallait trouver un vrai but utile à la vie malgré nos incertitudes, on peut penser que ce serait celui-là.

## II. Critique de la notion de néant

La pensée critique, le scepticisme et les résultats de la science ont apparemment eu raison des conceptions religieuses et mystiques du monde chez nombre de nos contemporains. Cette réaction est bien davantage psychologique et sociologique que logique, car la science s'occupe des faits empiriques et seulement d'eux ; tel n'est donc pas son rôle que d'ériger son matérialisme méthodologique en métaphysique, et de l'imposer en seule métaphysique valable (contre les métaphysiques spiritualistes par exemple), même si certains, à commencer par des scientifiques, entretiennent, le plus souvent inconsciemment, la confusion. Il faut avouer cependant que nombre des personnes croyantes ou anciennement croyantes ne le sont pas par ferveur spirituelle mais parce qu'elles estiment ou estimaient que leur foi disposait d'une assise un tant soit peu, sinon scientifique, du moins empirique et raisonnable. Que l'on mette à mal cette assise, et c'est la majestueuse cathédrale des doctrines religieuses et des croyances surnaturelles qui s'effondre. « Heureux ceux qui auront cru sans avoir vu ! » (Jean 20 : 29).

En fait, il serait sans doute plus juste de dire que si nombre de nos contemporains ont bel et bien évacué de la religion les contenus doctrinaux concernant Dieu, l'âme, l'immortalité, etc., en revanche, il subsiste toujours quelque part, une certaine forme de sentiment religieux vide d'objet, d'aspiration spirituelle frustrée, ce qui tend à induire un certain malaise. Pour réduire ce malaise, cette coque évidée se remplit avec de nouvelles représentations, qui paraissent peut-être plus conformes à l'air du temps, mais qui n'en demeure pas moins illégitimes. C'est ce que l'on pourrait appeler des tentatives de resacralisation.

Par exemple, certains tenteront de justifier scientifiquement leur croyance en Dieu, en un au-delà, en l'existence d'un esprit séparé du corps. Mais de telles tentatives ne peuvent aboutir ou relèvent, semble-t-il, de l'imposture et de la pseudoscience. En effet, toutes ces réalités supposées ne peuvent être appréhendées par une démarche scientifique rigoureuse qui passe nécessairement par la formulation claire des hypothèses, par la déduction de leurs conséquences empiriques attendues, et par la vérification de celles-ci. Or, si on prend par exemple le cas de Dieu, il n'existe pas de définition complète, consensuelle et rigoureuse de ce dernier, et les meilleurs éléments de définition dont nous disposions par la plume des grands théologiens n'ont pas de conséquences testables. Cela ne veut pas dire qu'il n'y a pas de Dieu, mais que s'il y en a un, la démarche scientifique n'est pas en mesure de le montrer, car elle s'occupe uniquement de faits empiriques. Certains suggèrent parfois que la science devrait évoluer, changer ses principes de manière à pouvoir évaluer la question, par exemple, de l'existence de Dieu. Mais les changements suggérés l'aliénerait, la rendrait inopérante à répondre aux questions sur lesquelles elle se penche actuellement avec succès. Ainsi, admettre des hypothèses infalsifiables en science, conduirait à traiter sur un plan d'égalité la théorie newtonienne de la chute des corps et la théorie selon laquelle c'est une fée invisible qui pousse les corps vers le bas… Pourquoi ne pas plutôt créer sa propre méthode d'investigation plutôt que de vouloir à tout prix redéfinir la science à sa façon ? C'est là qu'entrent en jeu, probablement, les facteurs d'ordre psychologique et sociologique : la science jouit d'un prestige que n'a pas n'importe quel système de pensée occulte. Aussi, l'espoir vain de persuader la communauté scientifique que la bonne façon de faire de la science est la sienne est un bon moyen, non seulement d'assurer une diffusion et une défense solide de ses idées, mais aussi et surtout, de réussir à se convaincre soi-même de leur justesse ! Mais la pseudoscience n'est pas seulement l'apanage des spiritualistes en quête de scientificité…

### *A. L'approche positive du néant (le néant comme état durant lequel nous n'expérimentons rien) et l'argument logique : un néant positif est logiquement contradictoire.*

Dans le camp matérialiste, on trouve également des dérives pseudoscientifiques (que l'on qualifiera plus volontiers de scientistes), mais plus subtiles, et moins critiquables d'un certain point de vue, puisque somme toute, si le matérialisme métaphysique n'est point requit par le matérialisme méthodologique, en revanche, la vérité du premier est la façon la plus simple de garantir la validité du second. Le scientisme affirmera que la science ayant montré que la conscience est liée au cerveau, à la mort de celui-ci, la conscience doit disparaître, oubliant au passage que la conscience est un phénomène à la fois objectif, certes, mais aussi subjectif. Cela ne pose pas de problème particulier de supposer la disparition de la conscience tant que l'on observe cela objectivement, en troisième personne. Toutefois, si l'on se demande ce que cela fait, pour la personne qui meurt, de mourir, puis d'être mort, le paradoxe surgit. Cela ne peut rien faire d'être mort puisque la mort est l'abolition de la conscience. Mais si cela ne fait rien d'être mort, en quoi le fait que je sois en vie se distingue-t-il du fait que je sois mort mais que quelqu'un me ressemblant comme deux gouttes d'eau ait pris ma place ? Toute notre vie, on expérimente le monde ou soi-même. Certes, le contenu précis de cette expérience change sans cesse, mais il semble qu'il reste toujours comme un écran sur lequel viennent se dessiner nos impressions, écran qui nous est tellement intime qu'on n'arrive pas à comprendre ce qu'il se passerait s'il venait à disparaître, ni même si une telle chose est vraiment possible en fin de compte. Cet « écran », c'est ce dont nous avons parlé comme étant le « soi témoin ».

Dans *L'Evolution Créatrice*, Bergson se livre, en vain, on l'a vu (cf. p.454 et 457), à cet exercice d'annihilation radicale :

> « Je ne me vois anéanti que si, par un acte positif, encore qu'involontaire et inconscient, je me suis déjà ressuscité moi-même. Ainsi j'ai beau faire, je perçois toujours quelque chose, soit du dehors, soit du dedans. Quand je ne connais plus rien des objets extérieurs, c'est que je me réfugie dans la conscience que j'ai de moi-même ; si j'abolis cet intérieur, son abolition même devient un objet pour un moi imaginaire qui, cette fois, perçoit comme un objet extérieur le moi qui disparaît. Extérieur ou intérieur, il y a donc toujours un objet que mon imagination se représente. Elle peut, il est vrai, aller de l'un à l'autre, et, tour à tour, imaginer un néant de perception externe ou un néant de perception intérieure, -mais non pas les deux à la fois, car l'absence de l'un consiste, au fond, dans la présence exclusive de l'autre. Mais, de ce que deux néants relatifs sont imaginables tour à tour, on conclut à tort qu'ils sont imaginables ensemble : conclusion dont l'absurdité devrait sauter aux yeux, puisqu'on ne saurait imaginer un néant sans s'apercevoir, au moins confusément, qu'on l'imagine, c'est-à-dire qu'on agit, qu'on pense, et que quelque chose, par conséquent, subsiste encore. » [300]

Mais en fin de compte, peut-être que cet écran intime résiduel est tout ce que l'on doit entendre par néant ? Selon cette idée, nous pourrions caractériser le néant, tout simplement comme le vide, ce qui reste, une fois que l'on a tout enlevé de notre champ de conscience. L'être pur, non déterminé. Dès lors, on pourrait l'assimiler à l'ensemble vide en mathématique. Or, l'ensemble vide étant constitutif de tous les ensembles, le néant serait constitutif de tout être, et donc de toute forme de vie. Par conséquent, loin d'être une étrangeté paradoxale, le néant serait ce qui nous est le plus intime, ce que, d'une certaine façon, nous connaissons le mieux (et qu'en même temps, nous avons de commun avec tout être). Il s'ensuivrait que l'état post-mortem – ainsi que ce que l'on pourrait appeler l'état

[300] BERGSON, Henri. *L'évolution créatrice*, Chap. IV.

« ante-embryonnaire » – serait la chose au monde la plus facile à s'imaginer, et même qu'on ne pourrait pas faire autrement que de se le représenter constamment, même si voilé – durant la vie – par toutes les autres représentations, sensations, images mentales dont il est le substrat inamissible, le dénominateur commun nécessaire, la condition de possibilité. Ainsi, pour Hegel, le néant « est égalité simple avec lui-même, vacuité parfaite, absence de détermination et de contenu, état de non-différenciation en lui-même ». Il est « la même détermination, ou plutôt la même absence de détermination, et partant absolument la même chose que l'être pur »[301].

Peut-être, mais cela ne semble pas être l'avis de la plupart des nihilistes. Tout d'abord, un tel néant n'a rien de définitif. De l'être pur indéterminé émerge constamment des êtres « impurs » et déterminés. Du reste, le terme de « néant » conviendrait-il vraiment pour désigner quelque chose, quoi que ce soit, qui reste et demeure ? Bergson poursuit :

> « L'image proprement dite d'une suppression de tout n'est donc jamais formée par la pensée. L'effort par lequel nous tendons à créer cette image aboutit simplement à nous faire osciller entre la vision d'une réalité extérieure et celle d'une réalité interne. Dans ce va-et-vient de notre esprit entre le dehors et le dedans, il y a un point, situé à égale distance des deux, où il nous semble que nous n'apercevons plus l'un et que nous n'apercevons pas encore l'autre : c'est là que se forme l'image du néant. En réalité, nous apercevons alors l'un et l'autre, étant arrivés au point où les deux termes sont mitoyens, et l'image du néant, ainsi définie, est une image pleine de choses, une image qui renferme à la fois celle du sujet et celle de l'objet, avec en plus, un saut perpétuel de l'une à l'autre et le refus de jamais se poser définitivement sur l'une d'elles, Il est évident que ce n'est pas ce néant-là que nous pourrions opposer à l'être, et mettre avant lui ou au-dessous de lui, puisqu'il renferme déjà l'existence en général. »

---

301 HEGEL, Georg Wilhelm Friedrich. *Science de la Logique*, I, 1

Le néant va donc être plutôt « conçu » comme l'absence ultime d'être, la parfaite inexistence, le non-être, et par là même comme l'absolument incompréhensible, inconnaissable et inimaginable, puisque le fait même d'être nous empêche de comprendre, connaître et imaginer le non-être. Il demeure cependant que ce néant est censé paradoxalement avoir une certaine valeur de positivité puisqu'il faut bien qu'il fasse la différence d'avec l'être, puisqu'il est notre lot dernier et définitif à tous. Il y a donc une tension qui se manifeste entre une conception du néant comme non-être et par l'idée, simultanée, selon laquelle ce non-être est ce vers quoi nous tendons.

Pour palier cet incompréhensible ultime, l'imagination prend le relais et, de même qu'elle peuplait autrefois le Ciel de créatures surnaturelles, elle se contorsionne aujourd'hui et se malmène pour tenter de dépeindre le néant de la mort, pour décrire au mieux l'indescriptible. Sa tâche, vaine et sans fin, est une perpétuelle reconstruction, un perpétuel recommencement digne du supplice de Sisyphe, puisque aucune image ne parvient à exprimer et à isoler au mieux la quintessence de l'idée de néant sans réintroduire subrepticement, au prix parfois de contours et de détours prodigieux, l'idée d'être. Les meilleures tentatives rejoignent en fait l'idée, esquissée plus haut, d'écran intime résiduel, mais alors ce n'est plus véritablement du néant tel que l'on s'efforce de le définir. De fait, la mort est souvent conçue, même par des esprits brillants (Asimov par exemple), comme une sorte de trou plus noir que le noir le plus noir dans lequel on tombe, après avoir rendu notre dernier souffle. Mais un trou d'une qualité très particulière, puisque dire qu'il est noir n'est encore qu'une approximation. Le néant, on l'obtient, non seulement en faisant le vide, mais en enlevant le vide lui-même ! Mais cela n'est point imaginable…

Que vaut cette pseudo-représentation « positive » du néant ? Le néant est-il quelque chose qui nous attend après la mort ? Probablement pas, car une telle conception du néant comme un état est tout bonnement contradictoire dans les termes. Il n'y a pas une chose, le néant, telle que cette chose a pour propriété de ne pas exister ! Par conséquent, concevoir le néant comme un trou noir, un sombre abysse, le royaume de l'oubli, etc. est totalement hors de propos.

### *B. L'approche négative du néant (le néant n'est pas un état) et l'argument sémantique/non-cognitiviste : le concept d'annihilation est un concept sémantiquement vide.*

Une deuxième approche tentera de concevoir le néant non pas comme un état, mais comme une absence d'état. Ainsi, il n'y aurait pas lieu de se demander quel effet cela fait d'être mort, car être mort n'est pas quelque chose que l'on expérimente, et par conséquent, il n'y a pas d'effet que cela fait d'être mort.

Une telle conception, pour raisonnable qu'elle semble être, ne tient pas la route pour autant. En effet, qu'est-ce qu'une absence d'état ? On peut caractériser l'absence d'un état particulier, mais c'est toujours relativement à un autre ou à d'autres.

On ne peut ni imaginer ni concevoir le concept d'une pure négation d'être. Citons à nouveau Bergson[302], qui a très bien analysé cela :

---

[302] BERGSON, Henri. *L'évolution créatrice*, Chap. IV.

« En fait, l'objet qu'on supprime est ou extérieur ou intérieur : c'est une chose ou c'est un état de conscience. Considérons le premier cas. J'abolis par la pensée un objet extérieur : à l'endroit où il était, « il n'y a plus rien ». - Plus rien de cet objet, sans aucun doute, mais un autre objet a pris sa place : il n'y a pas de vide absolu dans la nature. Admettons pourtant que le vide absolu soit possible ; ce n'est pas à ce vide que je pense quand je dis que l'objet, une fois aboli, laisse sa place inoccupée, car il s'agit par hypothèse d'une *place,* c'est-à-dire d'un vide limité par des contours précis, c'est-à-dire d'une espèce de *chose.* Le vide dont je parle n'est donc, au fond, que l'absence de tel objet déterminé, lequel était ici d'abord, se trouve maintenant ailleurs et, en tant qu'il n'est plus à son ancien lieu, laisse derrière lui, pour ainsi dire, le vide de lui-même. Un être qui ne serait pas doué de mémoire ou de prévision ne prononcerait jamais ici les mots de « vide » ou de « néant » ; il exprimerait simplement ce qui est et ce qu'il perçoit; or, ce qui est et ce qu'on perçoit, c'est la *présence* d'une chose ou d'une autre, jamais *l'absence* de quoi que ce soit. Il n'y a d'absence que pour un être capable de souvenir et d'attente. Il se souvenait d'un objet et s'attendait peut-être à le rencontrer : il en trouve un autre, et il exprime la déception de son attente, née elle-même du souvenir, en disant qu'il ne trouve plus rien, qu'il se heurte au néant. Même s'il ne s'attendait pas à rencontrer l'objet, c'est une attente possible de cet objet, c'est encore la déception de son attente éventuelle, qu'il traduit en disant que l'objet n'est plus où il était. Ce qu'il perçoit, en réalité, ce qu'il réussit à penser effectivement, c'est la présence de l'ancien objet à une nouvelle place ou celle d'un nouvel objet à l'ancienne; le reste, tout ce qui s'exprime négativement par des mots tels que le néant ou le vide, n'est pas tant pensée qu'affection, ou, pour parler plus exactement, coloration affective de la pensée. L'idée d'abolition ou de néant partiel se forme donc ici au cours de la substitution d'une chose à une autre, dès que cette substitution est pensée par un esprit qui préférerait maintenir l'ancienne chose à la place de la nouvelle ou qui conçoit tout au moins cette préférence comme possible. Elle implique du côté subjectif une préférence, du côté objectif une substitution, et n'est point autre chose qu'une combinaison, ou plutôt une interférence, entre ce sentiment de préférence et cette idée de substitution.

Tel est le mécanisme de l'opération par laquelle notre esprit abolit un objet et arrive à se représenter, dans le monde extérieur, un néant partiel. Voyons maintenant comment il se le représente à l'intérieur de lui-même. Ce que nous constatons en nous, ce sont encore des phénomènes qui se produisent, et non pas, évidemment, des

> phénomènes qui ne se produisent pas. J'éprouve une sensation ou une émotion, je conçois une idée, je prends une résolution : ma conscience perçoit ces faits qui sont autant de présences, et il n'y a pas de moment où des faits de ce genre ne me soient présents. Je puis sans doute interrompre, par la pensée, le cours de ma vie intérieure, supposer que je dors sans rêve ou que j'ai cessé d'exister ; mais, à l'instant même où je fais cette supposition, je me conçois, je m'imagine veillant sur mon sommeil ou survivant à mon anéantissement, et je ne renonce à me percevoir du dedans que pour me réfugier dans la perception extérieure de moi-même. C'est dire qu'ici encore le plein succède toujours au plein, et qu'une intelligence qui ne serait qu'intelligence, qui n'aurait ni regret ni désir, qui réglerait son mouvement sur le mouvement de son objet, ne concevrait même pas une absence ou un vide. La conception d'un vide naît ici quand la conscience, retardant sur elle-même, reste attachée au souvenir d'un état ancien alors qu'un autre état est déjà présent. Elle n'est qu'une comparaison entre ce qui est et ce qui pourrait ou devrait être, entre du plein et du plein. En un mot, qu'il s'agisse d'un vide de matière ou d'un vide de conscience, la représentation du vide est toujours une représentation pleine, qui se résout à l'analyse en deux éléments positifs ; l'idée, distincte ou confuse, d'une substitution, et le sentiment, éprouvé ou imaginé, d'un désir ou d'un regret. »

### *C. Conclusion*

Nous pouvons présenter un argument général contre l'idée que la mort conduit au néant.

1. Lorsque une entité (quelle qu'elle soit : chose, processus, propriété, etc.) change, il y a toujours, de proche en proche dans le temps, quelque chose qui demeure, et par rapport à quoi l'on peut parler de changement, et non de simple substitution (par exemple, une table repeinte en bleu, un homme qui vieilli, etc.)

2. Lorsqu'une entité (quelle qu'elle soit : chose, processus, propriété, etc.) cesse d'exister, il

y a toujours un contexte par rapport auquel elle cesse d'exister, ce qui permet de caractériser positivement cette disparition (ce n'est jamais tout le réel qui disparaît, et si toutes les entités qui constituent le réel disparaissaient, alors ce qui resterait serait le réel)

3. De 1. et 2. nous pouvons suggérer qu'une disparition d'une entité est une certaine sorte de changement qui affecte le réel dans sa totalité.

4. Tout état doit pouvoir être caractérisé positivement par un ensemble de propriétés

5. Ou bien le néant expérientiel de la mort (et non le néant de la mort en troisième personne) est un état ou bien il n'en est pas un.

    5.1. S'il n'est pas un état, alors il n'existe pas de changements affectant l'expérience individuelle qui puisse y conduire car entre la dernière étape conduisant au néant et le néant lui-même, il n'y aurait rien qui demeure (le néant ne peut pas être considéré comme une telle chose qui demeure).

    5.2. S'il est un état, alors on peut le définir positivement :

        5.2.1. Ou bien par rapport à quelque chose que sa survenue modifie dans la conscience expérientielle.

        5.2.2. Ou bien par rapport à quelque chose que sa survenue modifie dans le réel.

6. On ne peut caractériser positivement ce que le néant modifierait dans la conscience

expérientielle (dire qu'elle la ferait disparaître n'est pas une caractérisation positive mais une caractérisation négative) ou dans le réel (la conscience expérientielle n'est pas une partie du monde réel considéré comme monde objectif).

7. De 5.1 et de 6 on en conclut que la conscience expérientielle ne peut disparaître.

NB : Un raisonnement semblable permettrait de montrer que la conscience expérientielle ne peut pas non plus apparaître, c'est-à-dire avoir de commencement dans le temps.

Bien sûr, il y a une objection toujours possible à nos tentatives de réfutation de la notion de néant. On peut supposer que le néant « existe » (avec un sens du mot « exister » transcendant ce qui est compréhensible par nos facultés cognitives) quand même, mais qu'il échappe à notre compréhension. Un néant métaphysique. On pourrait soutenir que, si on ne peut pas imaginer sa propre inexistence en première personne, cela relève d'une limite cognitive, et non nécessairement d'une impossibilité métaphysique. Ainsi, le fait qu'un daltonien ne verrait que du noir, du blanc et des niveaux de gris, n'implique pas qu'il n'existe pas d'autres couleurs. Elles sont seulement inaccessibles à son système perceptif.

Bien sûr, il n'est pas question pour nous de nier qu'il puisse « exister » des choses par nature inaccessibles à notre entendement. Mais, pour la même raison, on ne peut non plus affirmer l'existence de ces choses. Une saine suspension de jugement est donc requise, en ce qui concerne le néant métaphysique. Encore que l'on puisse se demander l'utilité d'avancer une telle hypothèse qui ne va bien sûr pas de soi. Pourquoi ne pas également supposer l'existence de cercles carrés, à ce compte-là ?

On peut également faire remarquer que, par rapport au cas des couleurs dont notre daltonien ne parvient pas à se faire une conception positive, le cas du néant dont on ne peut également se faire une conception positive, diffère en plusieurs points :

- L'incapacité du daltonien à se représenter les couleurs est restreinte à son cas et à quelques autres (d'autres daltoniens). L'incapacité de se représenter le néant subjectif est générale. Personne ne peut se le représenter. Alors que de nombreuses personnes pourront témoigner auprès de notre daltonien de l'existence d'autres couleurs que le noir, le blanc et le gris, personne ne pourra témoigner de l'« existence » du néant, car personne ne peut savoir, ou expérimenter le néant. Le néant est inconcevable, non seulement positivement, mais même négativement, en ce que le concevoir, revient à poser l'être qui s'y oppose. Il y a donc contradiction.

- Notre daltonien, bien que ne parvenant pas à se représenter les couleurs, peut cependant y faire référence. Il peut se dire : « Le rouge est ce que Untel expérimente lorsqu'il regarde une tomate bien mûre ». Rien de tel avec le néant. On pourrait se dire : « Le néant est ce qu'"expérimente" Untel après que son cerveau ait cessé définitivement de fonctionner » ou « C'est ce qui "attend" Untel après sa mort », mais à proprement parler, le néant ne peut rien attendre sans se nier lui-même, il ne peut être expérimenté sans cessé d'être le néant.

- Rechercher le néant c'est un peu comme rechercher ce qu'il y a au Nord du Pôle Nord. Par définition, et non seulement contingentement, c'est impossible. Alors que l'impossibilité qu'éprouvent certaines personnes à expérimenter les couleurs

n'est que contingente. Il « suffirait » de modifier leurs yeux et/ou leur cerveau pour pallier à cette difficulté, et cette modification peut, elle, être caractérisée positivement.

A vrai dire, si la vie débouche, à la mort, sur le néant, cela ne peut vouloir dire qu'une seule chose, c'est qu'elle ne débouche tout simplement pas, et que la mort est, pour nous, en première personne, un horizon dont on se rapproche continuellement sans jamais l'atteindre. De la même façon dont nous n'avons jamais débouché sur la vie, à la naissance, à partir du néant. Nous ne nous sommes jamais dit, surpris, « Tiens donc, c'est inédit, voilà que j'existe, voici que je vis mes premières secondes de vie ! Jusqu'à présent, je n'avais jamais existé ! »[303]

Nous conclurons donc avec Bergson :

> « Il suit de cette double analyse que l'idée du néant absolu, entendu au sens d'une abolition de tout, est une idée destructive d'elle-même, une pseudo-idée, un simple mot. Si supprimer une chose consiste à la remplacer par une autre, si penser l'absence d'une chose n'est possible que par la représentation plus ou moins explicite de la présence de quelque autre chose, enfin si abolition signifie d'abord substitution, l'idée d'une "abolition de tout" est aussi absurde que celle d'un cercle carré. L'absurdité ne saute pas aux yeux, par ce qu'il n'existe pas d'objet particulier qu'on ne puisse supposer aboli : alors, de ce qu'il n'est pas interdit de supprimer par la pensée chaque chose tour à tour, on conclut qu'il est possible de les supposer supprimées toutes ensemble. On ne voit pas que supprimer chaque chose tour à tour, consiste précisément à la remplacer au fur et à mesure par une autre, et que dès lors la suppression de tout absolument implique une véritable contradiction

[303] Et nous ne pourrions pas nous le dire (nous voulons dire : sérieusement et sincèrement, pas juste comme plaisanterie), quand bien même nous serions venus au monde avec toutes les capacités cognitives et linguistiques d'un adulte.

> dans les termes, puisque cette opération consisterait à détruire la condition même qui lui permet de s'effectuer. »

# III. Arguments pour l'éternité

## *A. Il était a priori très improbable que j'existe maintenant*

Voici quelques idées issues de ce que l'on pourrait appeler le « sens commun moderne » :

1. Le temps s'écoule indépendamment de moi. Avant que je naisse, le temps s'écoulait déjà, et lorsque je serai mort, le temps continuera de s'écouler.

2. Je suis mon cerveau ou tout au moins, je suis existentiellement lié à lui ; lorsque mon cerveau meurt, je meurs

3. L'identité : Je suis le même « Je » depuis ma naissance jusqu'à ma mort, en dépit des changements qui peuvent m'affecter. En revanche, je ne suis pas les autres, je suis un être singulier, séparé, indépendant ontologiquement des autres. A la limite, il pourrait être concevable que je sois seul à exister et que tous les autres soient illusion.

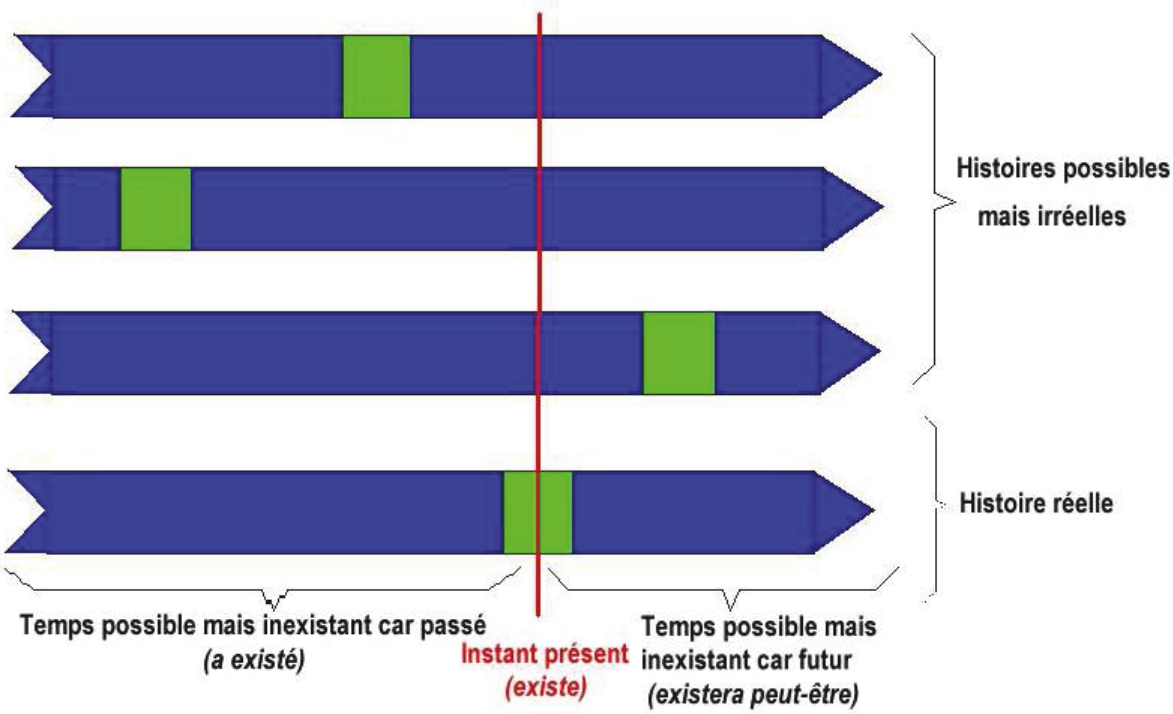
4. L'unicité de mon existence : Je n'ai pas plusieurs vies, ma vie est un intermède d'une centaine d'années entre deux néants d'infinies durées.

Enfin, un constat important : **J'existe maintenant.**

Les 4 premières idées rendent cependant miraculeux le constat de mon existence, maintenant.

En effet, le temps s'écoule depuis 15 milliards d'années au moins, et en a encore pour un bon bout de temps à s'écouler. Pour fixer les choses, disons 15 milliards d'années encore. En tout, cela fait 30 milliards d'années. Une vie humaine dure (pour simplifier) 100 ans. Je devrais donc déjà être mort ou pas encore né avec une probabilité de 29 999 999 900/30 000 000 000 = 0,99999999666666666666666666666667 !!! Même en ne considérant que les 100 000 ans de l'histoire humaine depuis l'homme de Cro-Magnon, la probabilité reste de 0,999 !

Nous pouvons récapituler ce raisonnement sur le schéma suivant, pour plus de clarté :




**Flèche du temps**

**En vert : Fraction de temps où j'existe.**
**En bleu : Fraction de temps où je n'existe pas**

Il semble donc que pour rendre mon existence maintenant moins miraculeuse nous devions opter pour une au moins de ces options :

1. Le temps s'écoule subjectivement, si bien que tous les moments du temps sont présents, relativement à un certain référentiel. Cela a une conséquence importante : je ne peux atteindre un instant qui me ferait disparaître (car cela voudrait dire que l'écoulement du temps peut me faire sortir – et donc peut sortir lui-même – du référentiel par rapport auquel il existe). J'existe donc éternellement dans mon propre référentiel temps, qui, pourquoi pas, se rembobine pour une nouvelle lecture à l'identique à chaque fois que je meurs.

2. Je ne suis pas mon cerveau. Soit parce que j'en suis ontologiquement distinct (spiritualisme, dualisme), soit parce que mon identité et mon esprit dépendent ontologiquement, non seulement de mon cerveau, mais également de mon corps et de mon environnement physique et social (externalisme actif, mémétique, théorie du cerveau planétaire, etc.). Ainsi, lorsque mon cerveau meurt, il n'est pas nécessaire que je meurs également : je peux survivre sous forme d'une âme désincarnée (spiritualisme, dualisme), ou (littéralement) à travers mes proches, l'environnement, mes œuvres, ma descendance.

3. Soit l'identité personnelle est un leurre, lié à la clôture phénoménologique de la conscience ainsi qu'à la mémoire (théorie phénoménaliste), soit elle correspond à un ensemble de propriétés objectives et en principe reproductibles à d'autres moments du temps (théorie physicaliste).

4. Je n'ai pas qu'une seule vie, j'ai déjà vécu avant ma naissance, et je vivrai encore après ma mort (réincarnationnisme).

Il nous faut toutefois considérer une objection de taille : un phénomène, fut-il improbable, n'est pas impossible. L'importance subjective que nous accordons – pour ainsi dire par construction – à notre propre existence ne doit pas nous aveugler quant à son importance objective. De nombreux phénomènes improbables ont lieu tous les jours. Chaque tirage du Loto *effectivement* sorti avait extrêmement peu de chance, *a priori*, de sortir. Nul besoin, pourtant, de s'extasier du fait que les numéros qui sont sortis sont effectivement sortis !

Toutefois, le cas de l'identité personnelle est quelque peu différent, en ce sens qu'il semble bizarre de se dire que de nombreuses personnes possibles n'existeront jamais, et de se dire que tout cela ne se joue qu'à un poil près dans la programmation génétique des individus. On est tenté de se dire, d'une part, que si on avait été un tout petit peu différent, par exemple blond au lieu de brun, on aurait quand même existé, et, d'autre part, qu'une personne possible qui n'existerait nulle part et jamais ne serait tout simplement pas une personne possible. Contrairement au Loto, où l'on peut établir théoriquement quelles sont toutes les combinaisons possibles et imaginables, même si ces combinaisons n'étaient jamais réalisées, on ne peut pas, quant à l'identité personnelle, établir théoriquement quelles sont toutes les personnes possibles et imaginables. On a envie de se dire que les seules personnes possibles sont celles qui en viennent à exister effectivement à un moment donné, ou encore, et cela revient pratiquement au même, qu'aucune personne particulière n'est théoriquement possible, bien qu'elle existe, existait ou existera. Il semble donc que, dans le premier cas, les possibilités d'existence des personnes, dans le second cas, les existences effectives des personnes, tiennent bien du « miracle ».

### *B. Ce qui ne dure pas ne peut pas être définitif*

On considère habituellement le sommeil profond, sans rêve, comme une bonne métaphore du néant de la mort. A ceux qui ne peuvent concevoir le néant ou que celui-ci effraie, on fera remarquer que la mort est comme un long sommeil dont on ne se réveille pas, et qu'il n'y a pas plus de raison de la craindre que d'avoir peur de s'endormir la nuit.

Le sommeil profond serait donc comme une sorte de mort ou de néant provisoire, dont on revient chaque fois que l'on se réveille. Que pensez de cette interprétation ?

La conscience s'éteint provisoirement, semble-t-il, durant le sommeil profond. Ceci, d'après certains, devrait nous aider à mieux comprendre ce qui se passe après la mort. Seulement voilà, au réveil, tout ce dont on se souvient ce sont : de vagues impressions et pensées (cela n'est assurément pas ça, le néant), des rêves (cela n'est pas ça non plus) et l'impression que quelques heures se sont écoulées en quelques minutes. Le néant du sommeil, c'est plusieurs heures dans la perception habituelle de la durée qui disparaissent soudainement, s'écoulent en un instant. Le néant, c'est un voyage temporel de la conscience vers l'avenir. Mais ceci n'est constatable que lorsque l'on est à nouveau conscient ! Et encore, cela n'est constatable que pour autant que l'on ait une montre, le soleil ou des congénères qui nous indiquent quelle heure il est. Sans cela, les phases de néant de notre conscience seraient tout bonnement indétectables. Imaginons le scénario suivant : toutes les secondes, l'univers entier, et toutes les consciences qu'il contient, disparaissent pendant 100 milliards d'années, puis réapparaissent, exactement dans l'état dans lesquels ils étaient, pendant une seconde, et ainsi de suite. Ca paraît terrifiant ! Pourtant, il n'y aurait aucun moyen de le savoir, aucun moyen de s'en rendre compte car le néant n'a pas de durée. D'ailleurs, dans l'absolu, qui nous dit que ce n'est pas ainsi que l'univers fonctionne ? Finalement, ce qui est difficile à appréhender, dans l'idée du néant après la mort, c'est que quelque chose, **pour nous de durée nulle**, puisse sceller notre sort **éternellement**. Quelque soit la durée finie de temps objectif durant laquelle ma conscience est interrompue, lorsqu'elle reprend, c'est, de son point de vue, comme s'il ne s'était rien passé de spécial. Le dernier phénomène psychique précédant l'interruption est suivi directement par le premier phénomène psychique la suivant, dans une parfaite continuité. Le néant est donc sans durée subjective. Pourtant, la mort est censée nous anéantir définitivement, tant objectivement que subjectivement.

Si mon vélo est définitivement hors d'usage, cela implique un certain changement, affectant mon vélo, qui dure.

Si ma subjectivité est définitivement détruite, cela implique un certain changement, affectant ma subjectivité, qui dure.

Or le néant ne durant pas, du point de vue de ma subjectivité, celle-ci ne pourrait donc être affectée durablement par celui-là.

### *C. Le temps ne s'écoule pas réellement*

Un autre argument en faveur de l'impossibilité du néant après la mort est que les distinctions entre le passé, le présent et l'avenir sont illusoires. Ce que l'on peut dire, c'est que tel événement du temps est avant, pendant ou après un autre, mais pas que cet événement est passé, présent ou futur. Si ces distinctions sont illusoires, alors l'écoulement du temps est illusoire. Et si l'écoulement du temps est illusoire, alors on ne peut mourir car en aucun cas un écoulement du temps qui ne dépende que de mon esprit ne pourrait m'amener à une position temporelle où cet esprit n'existerait plus.

McTaggart[304] appelle « série A » les qualificatifs du temps tels que « passé », « présent », « futur ». Ces qualificatifs ne sont pas permanents. Ce qui est passé a été présent et futur, ce qui est présent a été futur et sera passé, ce qui est futur sera présent puis passé.

---

304 MCTAGGART, John McTaggart Ellis, "The Unreality of Time". *In:* LE POIDEVIN, Robin et MCBEATH, Murray. *The Philosophy of Time.*

Il appelle « série B » les qualificatifs temporels tels que « avant », « pendant », « après ». Ces qualificatifs sont permanents. Il est éternellement vrai que la seconde guerre mondiale est située temporellement après la première guerre mondiale, par exemple.

Pour McTaggart, les propriétés des séries A conduisent à une contradiction, car les séries A sont mutuellement exclusives. En effet, un événement ne peut être à la fois passé, présent et futur. En effet, soit une position temporelle donnée M. Cette position est présente, a été future et sera passée. Mais « sera passée » est équivalent à « est passée à une position temporelle future » et « a été future » est équivalent à « est future à une position temporelle passée ». Le futur est donc défini par rapport au passé et le passé par rapport au futur. Ces définitions circulaires montrent l'incohérence des séries A et donc, au moins une certaine irréalité du temps.

Le philosophe contemporain Francis Kaplan défend une conception similaire. Dans « L'irréalité de la mort » [305], il commence par passer en revue une série d'arguments issus de la physique, notamment la théorie de la relativité et la mécanique quantique, qui tendent à montrer l'irréalité du temps (ou plutôt, faudrait-il dire, le caractère illusoire de la distinction passé, présent, futur).

Ensuite, l'auteur présente des arguments plus psychologiques.

Tout d'abord, il nous montre qu'il existe un « présent du passé » :

---

[305] KAPLAN, Francis. « L'irréalité de la mort », *In* : LENOIR, Frédéric et DE TONNAC, Jean-Philippe, (Eds.). *La Mort et l'Immortalité. Encyclopédie des Savoirs et des Croyances*, p. 1434-1435.

> « Lorsque j'entends le dernier mot d'une phrase ou la dernière note d'une mélodie, j'ai présent à l'esprit les notes ou les mots précédents, sinon je n'aurais pas entendu une phrase ou une mélodie, mais une note ou un mot ou même seulement une partie d'un mot, il y aurait eu succession de sensations et non sensation d'une succession. Mais cela signifie que le début de la durée, les notes, les mots, les sons précédents qui, parce qu'ils sont précédents, sont passés, sont en même temps présents. »

Il montre ensuite que l'on ne peut entièrement rendre compte de ce « présent du passé » en faisant appel à la mémoire, car :

> « [s]ans doute, le souvenir, au moment où on en prend conscience, est une réalité présente […]. Il n'y a donc rien d'étonnant à ce que toutes les notes passées de la mélodie ou tous les mots passés de la phrase que j'entends soient présents au moment où j'entends le dernier mot de la phrase ou la dernière note de la mélodie. Mais ou je prends conscience successivement de chacune des réalités présentes correspondant à chacune des notes passées de la mélodie, à chacun des mots passés de la phrase et nous nous retrouvons devant le même problème : ces prises de conscience ne sont pas présentes en même temps et il faut une nouvelle mémoire pour me rappeler, au moment de la prise de conscience du souvenir de la dernière note passée ou du dernier mot passé, des prises de conscience des souvenirs des notes ou des mots précédents — en quelque sorte des souvenirs de souvenirs et évidemment des souvenirs de souvenirs de souvenirs et ceci indéfiniment. Ou je prends conscience en même temps, au même instant, de chacune des réalités présentes correspondant à chacune des notes ou des mots passés et elles se télescoperaient, se confondraient; ce ne serait plus une phrase, mais une cacophonie; ce ne serait plus une mélodie, mais dans la meilleure des hypothèse, ce qu'on appelle un accord musical. »

Il reste à comprendre l'ordre de succession dont nous faisons l'expérience. En effet, si nous expérimentons toutes les notes d'une mélodie au présent, comment pouvons-nous distinguer l'ordre de succession des notes ? Si nous nous remémorons un souvenir du

passé, comment savons-nous qu'il concerne le passé ? On pourrait supposer qu'il y a une différence de qualité en fonction de l'ancienneté d'une réalité expérimentée, mais il y a un problème car :

> « comment puis-je savoir que des qualités différentes correspondent à une succession temporelle si je n'ai pas constaté au moins une fois que telle qualité corresponde à telle ancienneté et telle autre qualité à telle autre ancienneté et cela suppose que j'ai l'expérience de l'ancienneté indépendamment de la réalité présente du souvenir, autrement dit, que mon souvenir ne se borne pas à sa réalité présente. »

Pour y répondre, Kaplan suggère que :

> « [d]ans la mémoire, nous ne sommes pas en rapport avec le souvenir, mais réellement avec le passé à travers le souvenir ; nous ne prenons pas conscience du souvenir, mais du passé au moyen du souvenir. Certes, je peux prendre conscience du souvenir, mais c'est par une conscience réflexive qui n'est pas la mémoire et qui suppose une conscience non réflexive préalable qui n'est pas conscience du souvenir, mais conscience du passé. »

On notera qu'une telle position n'est pas sans se rapprocher de ce que l'on pourrait appeler un *réalisme mnésique direct*. De même que le réalisme perceptif direct suppose une relation directe de la conscience perceptive avec son objet, sans passer par une image interne qui serait la seule à être directement expérimentée, le réalisme mnésique direct supposerait ainsi une relation directe de la conscience, à travers l'expérience du souvenir, avec un authentique événement du passé.

Cette conclusion à de quoi surprendre, mais, à partir du moment où l'on admet que les phénomènes physiques à l'origine de la conscience sont des phénomènes dynamiques et

non statiques, donc s'inscrivant dans la temporalité, alors que le souvenir soit une conscience directe du passé n'est pas si choquant.

Une telle vision des choses est indirectement appuyée par la défense de l'externalisme actif que nous avons esquissée dans le corps de cette thèse. De même que le réalisme perceptif direct trouve un appui à travers l'externalisme actif que nous pourrions appeler « spatial », le réalisme mnésique direct trouverait un appui à travers une autre forme d'externalisme actif que nous pourrions appeler « temporel » (la suggestion est personnelle et ne se trouve pas chez Kaplan). L'externalisme actif spatial considère que la conscience, émergeant sur la base d'une dynamique impliquant le corps, le cerveau, et l'environnement physique et social, transcende l'espace. L'externalisme actif temporel prolongerait l'idée en suggérant que la conscience émerge aussi d'une dynamique impliquant les diverses phases temporelles de ces entités. En effet :

> « puisque la conscience, par la mémoire, peut atteindre le passé, c'est qu'elle transcende le temps ; car si elle restait dans l'ordre du temps, le passé, pour elle, ne serait plus et, par conséquent, elle ne pourrait l'atteindre. »

## *D. Un argument général en faveur de l'éternité du sujet.*

Nous aimerions terminer ce parcours par l'évocation d'un dernier argument en faveur de l'éternité du sujet. L'avantage de cet argument est qu'il est très général, il ne dépend aucunement de notre conception de l'identité personnelle. Il implique de considérer la personne que nous sommes actuellement. Y a-t-il, au sein de cette personne, quelque chose, quoi que ce soit, qui vous paraît essentiel pour vous ? Quelque chose que vous tenez pour à la fois absolument indispensable et suffisant à votre survie ? Quelque chose qui,

d'après vous, doit exister demain pour qu'il soit vrai que vous existez demain ? Un réquisit minimum qui satisferait votre désir de survie ?

Admettons que ce quelque chose soit la mémoire. Il y a sans doute encore moyen d'affiner cette première réponse. Tous vos souvenirs sont-ils importants au même titre ? Probablement que non. Des connaissances générales sur le monde, des souvenirs d'enfance, d'événements marquants, un certain nombre de traits psychologiques, etc. sont sans doute plus importants que de savoir ce que vous avez mangé avant-hier. Si on vous amputait de tout ce que vous êtes mais que l'on conservait ces éléments précités, vous considéreriez sans doute que vous êtes encore vivant, que vous existez toujours.

Après votre mort, voir bien après, il n'est pas impossible, il est même très probable, que vienne au monde quelqu'un qui acquérra grosso modo les mêmes connaissances générales sur le monde que vous, qui aura une enfance assez similaire et développera une personnalité assez semblable, qui vivra des événements importants comparables à ceux que vous avez vécus : qui se mariera, aura des enfants, trouvera un travail, etc. Bien sûr : il ne se mariera pas avec la même personne que vous, ni n'aura les mêmes enfants ni nécessairement exactement le même travail, mais si votre vie avait été différente sur ces points, êtes-vous prêt à considérer que cela vous aurait anéanti (au sens propre) à ce point ? Cette personne satisfera donc les *desiderata* que vous avez posés pour que vous puissiez légitimement affirmer : « Je vis, j'existe ! ». Si elle existe, il existera ce qui est essentiel, selon vous, pour que vous existiez. Si vous avez raison au sujet de ce que vous avez déterminé être l'essentiel de vous-même, alors vous existerez si cette personne existe. Et il est probable, en effet, qu'une telle personne existe un jour.

Maintenant, peut-être voudrez-vous considérer un autre critère que la mémoire ? Admettons que vous teniez pour essentiel la matière qui vous constitue. Cette matière, ou la matière qui vient la remplacer dans un processus de renouvellement très progressif, n'est pas détruite par la mort. Et lorsque votre corps se dissout dans la Terre, le processus de renouvellement progressif se poursuit. Certain des éléments qui constituent votre cadavre sont drainés au loin par les micro-organismes, l'eau, etc. D'autres éléments sont apportés, dans un flux qui se continue bien au-delà de la mort et qui s'élargit, se diffuse. En effet, si votre corps constitue encore une unité relativement fermée sur elle-même, le processus de la décomposition fusionne cette unité à une unité bien plus vaste qui est la Terre elle-même. De celle-ci, d'autres organismes naissent, qui seront en continuité matérielle avec elle, et donc avec vous. Certains seront composés des mêmes éléments, numériquement parlant, qui vous constitue aujourd'hui. Encore une fois, les conditions essentielles, selon vous, à votre vie, et donc à votre survie par-delà la mort, sont réunies.

Et combiner les deux critères ne change rien à l'affaire. En rajouter un troisième, un quatrième, un cinquième, etc. non plus, car du moment que ces critères sont objectifs, ils sont reproductibles, et s'ils sont reproductibles, ils ont une certaine probabilité d'être réalisés une nouvelle fois. Or, lorsque l'on donne le temps au hasard pour agir, la probabilité qu'une combinaison, même improbable, d'un certain ensemble de caractéristiques, apparaisse chez un individu, tend asymptotiquement vers un. Bien entendu, plus on est exigeant en termes de ce qui est requis pour notre survie, plus il faut de temps pour espérer voir ce réquisit s'actualiser. Mais, malgré tout, puisque la probabilité de cet événement n'est pas strictement de 0, la probabilité qu'il réapparaisse un jour à l'identique, si on se donne un temps infini, est de 1.

Poincaré affirmait que la vie est un éclair de lumière entre deux nuits infinies (quand bien même nous n'avons pas d'exemples de phénomènes absolument uniques pour étayer cette position). Contre cette conception, il semble donc, au contraire, que nous l'avons montré par cet argument, que nous sommes condamnés à vivre éternellement.

## IV. Quelle identité personnelle pour quelle survie ?

Dans la conclusion de notre thèse, nous avons présenté un certain nombre de critères de l'identité personnelle, qui tentent de mieux répondre à certaines questions problématiques (concernant l'identité spécifique et numérique, l'identité diachronique et synchronique). Nous pouvons intégrer ces définitions éparses au sein de modèles généraux de compréhension de l'identité personnelle.

Ces modèles, nous souhaiterions les examiner, afin de montrer quelles implications ils peuvent avoir, au sujet de la question de la mort, de la survie et de l'éternité. A défaut, nous aimerions montrer si l'on peut, et comment l'on peut, les concilier avec les vues développées jusqu'ici dans ces annexes.

Trois principaux modèles se dégagent.

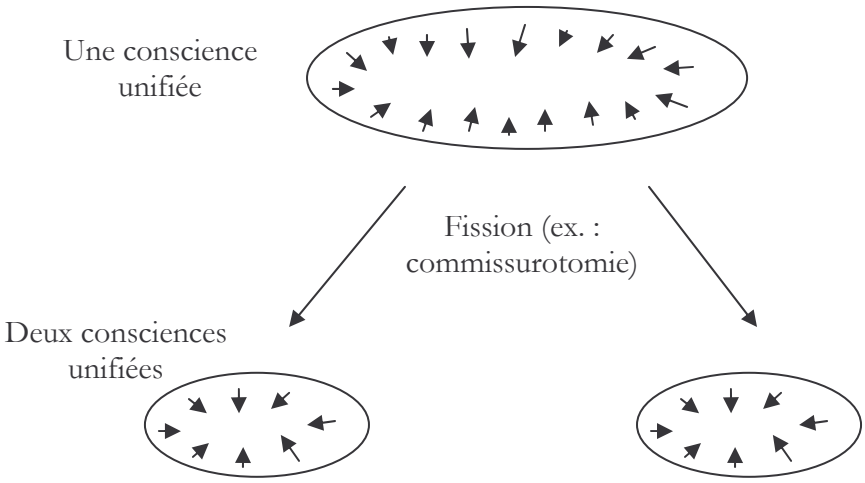
## *A. Premier modèle : théorie phénoménaliste de l'identité personnelle.*

Ce modèle s'inspire des travaux et réflexions, entre autres, de David Hume, Derek Parfit, Susan Blackmore, Thomas Nagel, Thomas W. Clarke, Wayne Stewart, et de la sagesse du Bouddhisme Theravada.

D'après ce modèle, il n'existe en réalité qu'une seule et unique personne dans tout l'univers. Cette personne, cependant, est très profondément dissociée. Chaque fragment étant phénoménologiquement clos sur lui-même, il a l'impression d'être un « moi » singulier, unique et indivisible. Cela n'est cependant qu'une illusion[306]. Selon ce modèle, la question de ce qu'il y avait avant la naissance ou de ce qu'il y aura après la mort n'a tout simplement pas de sens, car il n'y a pas d'ego réel. Or, de ce qui n'est pas réel, on ne peut pas dire qu'il commence ou fini d'exister.

On peut représenter ce modèle schématiquement de la sorte :

[306] A condition d'admettre – ce qui est problématique – que les sensations qui traversent notre champ d'expérience ont une identité propre, alors on peut admettre une identité – dérivée – de la personne, reposant sur une base phénoménaliste. Il faut suivre pour cela Leibniz qui voit dans une « continuation et liaison de perceptions » le fondement de l'identité réelle.




Les perspectives mystiques et eschatologiques de ce modèle sont un peu plus réjouissantes que ce que l'on pourrait penser de prime abord. Même si on ne peut parler d'immortalité concernant l'ego (mais on ne peut pas parler non plus d'annihilation), on peut entrevoir une possibilité de survie trans-personnelle, puisque finalement, la personne unique dont chaque « ego » ne serait qu'un fragment, n'est pas soumise, quant à elle, à la mort, du moins tant qu'il existe des personnes individuelles pour l'incarner.

Dans "Death, Nothingness and Subjectivity", Thomas W. Clark s'ingénie à nous rendre sensible cette intuition quant à la probabilité d'une forme de survie trans-personnelle selon une perspective phénoménaliste.[307]

[307] CLARK, Thomas W. "Death, Nothingness and Subjectivity".

Il est important de remarquer que dans la défense de son hypothèse, il reste dans un cadre strictement naturaliste, ce qui montre que, en dépit des idées reçues, point n'est besoin de faire intervenir des entités surnaturelles pour donner sens à une notion de survie au-delà de la mort individuelle.

Clark commence tout d'abord par critiquer la conception commune de la mort comme une plongée dans le néant, en s'attaquant, un peu à la manière de Bergson, à la cohérence d'un tel concept de « néant positif », arguant qu'il n'est pas peu contradictoire de soutenir qu'il existe quelque chose (le néant) qui n'existe pas.

Ensuite, il définie deux notions importantes pour la bonne compréhension de son article : la *continuité subjective personnelle* et le *sentiment subjectif d'avoir toujours été présent*.

La première notion ne doit pas être confondue avec la continuité psychologique, telle que la définit, par exemple, Parfit. Il s'agit de la continuité de conscience qui se manifeste entre deux périodes de conscience séparées par une période d'inconscience de la durée que l'on voudra. Du point de vue du sujet, tout se passe comme si la transition était instantanée. Il n'y a aucune expérience d'un néant de l'expérience au cours de notre vie consciente, et il ne peut, en fait, pas y en avoir, par définition, pour très précisément la même raison d'incohérence logique qui rend rationnellement inacceptable l'idée d'un plongeon dans le néant après notre mort.

La deuxième notion est une conséquence de la première, étendue aux larges périodes de temps qui précèdent notre naissance et suivent notre mort. *« Certainement,* nous dit Clark*, nous ne pensons pas ordinairement au temps avant notre venue au monde comme à un abysse duquel nous*

*avons réussi à échapper ; nous nous trouvons simplement présents au monde. Nous ne pouvons pas contraster le fait d'être conscient avec un état préalable de non-expérience. La même chose est vraie du temps après la mort. Il n'y aura aucun état personnel futur de non-expérience auquel nous pouvons comparer notre présent état d'être conscient. »*

Clark explique ensuite que malgré ce sentiment subjectif d'avoir toujours été présent, il reste difficile, lorsque la mort s'approche, de ne pas se demander « Que va-t-il m'arriver ? » ou « Qu'y a-t-il après ? ». Il est terriblement tentant de vouloir se projeter nous-même – ce *locus* d'expérience consciente – dans le futur, entrant dans la noirceur ou le vide de la non-expérience. Mais puisque le néant de la non-expérience à été disqualifié en tant que destinée de la subjectivité, quelle solution alternative proposer ?

Le fait que chaque conscience, pour elle-même, n'a ni début ni fin, ajouté au fait que lorsque l'on meurt, d'autres consciences nous survivent et nous succèdent, suggère une alternative à l'intuition d'un néant imminent à l'approche de la mort. Clark nous propose l'hypothèse suivante :

> « Au lieu d'anticiper le néant à la mort, je propose que nous devrions anticiper le *sentiment subjectif d'avoir toujours été présent*, expérimenté à travers un contexte différent, le contexte fourni par ces subjectivités qui existent ou vont exister.
>
> En proposant cela je ne cherche pas à suggérer qu'il existe des connexions surnaturelles et défiant la mort entre les consciences qui pourraient d'une quelconque façon préserver les éléments de mémoire ou de personnalité. Ce n'est pas du tout ce que j'ai en tête, puisque les preuves matérielles suggèrent que tout ce en quoi consiste une personne – un corps vivant, la conscience, la personnalité, les mémoires, les préférences, les attentes, etc. – sont effacées à la mort. La continuité subjective personnelle tel que je la définissais ci-dessus requiert que les expériences soient celle d'une personne *particulière* ; par conséquent, cette

> sorte de continuité est limitée par la mort. Donc quand je dis que *vous* devriez attendre avec impatience, à la mort, le « sentiment subjectif d'avoir toujours été présent », je parle plutôt vaguement, puisque ce n'est pas vous – pas *cet* ensemble de caractéristiques personnelles – qui va expérimenter le sentiment subjectif d'avoir toujours été présent. C'est plutôt un autre ensemble de caractéristiques (en fait d'innombrables ensembles) avec la capacité, peut-être, d'expérimenter différentes sortes de choses complètement différentes. Mais, en dépit de ces (peut-être radicales) différences, il partagera le qualitativement *très semblable sentiment* d'avoir toujours été ici, et, comme vous, n'expérimentera jamais sa cessation. »

Il semble que ce « sentiment subjectif d'avoir toujours été présent » puisse être rapproché de l'idée d'une personne unique dont nous sommes les fragments dissociés.

Clark fait appel à une expérience de pensée pour clarifier son propos. Il imagine qu'une machine, fruit d'une technologie avancée, nous permette un jour de maintenir en vie indéfiniment des personnes et de stopper et redémarrer à volonté leurs processus biologiques.

Imaginons une personne TC, caractérisée par un ensemble bien particulier de propriétés psychologiques, physiques, et de souvenirs, qui est installée dans cette machine.

On peut, à tout moment, la plonger dans l'inconscience (comme lors d'un sommeil profond) en arrêtant ses processus biologiques et la réveiller, après une durée quelconque. De son point de vue subjectif, ce gap temporel ne sera pas expérimenté, et tout se passera pour lui comme si sa dernière expérience, avant de sombrer dans l'inconscience, était immédiatement suivie par sa première expérience, à son réveil. Peut-être que des indices

extérieurs pourront lui apprendre qu'il a été inconscient pendant, par exemple un siècle, mais de son point de vue subjectif, il aura l'impression d'une transition instantanée.

Supposons à présent, que durant le long sommeil d'inconscience de TC, quelques légères modifications soient apportées à ses caractéristiques. Au réveil, on obtient TC-mod qui sera sans problème identifié par lui-même et son entourage en tant que TC, fut-ce sous une version modifiée, encore que cela peut passer inaperçu si les modifications sont très subtiles ou que des éléments de mémoire sont modifiés de manière à ce que TC-mod ait l'impression d'avoir toujours eu certaines caractéristiques pourtant nouvellement implantées. La subjectivité personnelle de TC a donc survécu.

Maintenant, supposons que durant l'état d'inconscience de TC, on opère à des modifications beaucoup plus importantes, radicales, de sa personnalité, de ses souvenirs, voire de son apparence physique. Il n'est pas dit que TC-rad, l'individu obtenu au réveil, soit encore la même personne que TC. Il ne se reconnaîtra probablement pas comme étant TC, et son entourage non plus. Pourtant, TC-rad, à son réveil, expérimente le sentiment d'avoir toujours été présent, qui est qualitativement le même que celui qu'expérimentait TC. En un certain sens, on peut donc dire que la première expérience de TC-rad suit donc immédiatement la dernière expérience de TC, du point de vue d'une subjectivité générique, qui est précisément définie par ce sentiment d'avoir toujours été présent.

Remarquons que si il semble clair que l'on peut identifier TC-mod à TC et affirmer en revanche, que dans le cas du passage de TC à TC-rad, TC cesse d'exister, et qu'une nouvelle personne, TC-rad apparaît, il n'est pas évident du tout de spécifier à partir de quel degré de changements apportés à TC, il convient de dire que TC cesse d'exister et que

quelqu'un d'autre vient à l'existence. Le flou de cette frontière est propre à appuyer l'intuition de Thomas W. Clark au sujet de la continuité, à travers TC-rad, du sentiment subjectif d'avoir toujours été présent, que ce dernier partage avec TC, et qui est un trait commun de toutes les subjectivités quelles qu'elles soient.

Appelons « mort par transformation » le processus faisant passer de TC à TC-rad. Thomas W. Clark affirme que la conscience est subjectivement continue, en un sens générique (le sentiment subjectif d'avoir toujours été présent), à travers une telle transformation. Considérée de « son » point de vue, l'expérience (au moins l'expérience subjective générique) ne stoppe jamais, bien qu'objectivement parlant (d'un point de vue extérieur) un certain contexte pour sa manifestation cesse et plus tard (aussi tard que vous pouvez l'imaginer) un autre contexte apparaît.

L'étape suivante de son argument consiste à appliquer cette conclusion à la mort et à la naissance (considérées dans leur sens ordinaire). Admettons à présent que TC, dans son sommeil, soit sujet à une mort cérébrale irréversible, à cause d'une erreur technique. Ensuite, un peu plus tard ou bien plus tard, quelque part dans le monde, un être humain naît (naturellement ou artificiellement). En dépit de l'absence de connexions causales entre TC et cet être humain, Thomas W. Clark pense que la situation est la même que pour la mort par transformation. C'est-à-dire qu'un contexte de conscience a pris fin et un autre, très différent, commence. Durant l'intervalle de temps objectif, il n'y a aucun hiatus subjectif dans la conscience ; seul le contexte de l'expérience a changé.

Pour plus de clarté et de plausibilité, récapitulons en quelques points l'argumentation de Thomas W. Clark :

1) C'est un fait communément admis que lorsque je meurs, les autres sujets continuent d'exister, par conséquent la subjectivité est très certainement immunisée contre *ma* mort dans *ces* circonstances.

2) Si je suis inconscient pendant quelque longueur de temps que ce soit, je n'expérimente pas cet intervalle ; je suis toujours « présent » ; c'est la *continuité subjective personnelle*.

3) Si, après une période d'inconscience, la personne transformée qui se réveille n'est pas moi, il n'y aura encore aucun trou perçu dans la conscience. La personne qui se réveille ressent, comme je le ressentais (donc ressent « encore »), qu'elle a toujours été présente. Il n'y a aucune expérience antérieure de n'avoir pas été présent, pas plus que je n'ai, en cessant d'exister, eu une telle expérience ; c'est la *continuité subjective générique*.

4) La mort et la naissance sont « fonctionnellement équivalent » à la sorte de transformation évoquée en 3), donc à nouveau il n'y aura pas de gap perçu, aucun néant de non-expérience dans lequel le sujet pourrait sombrer. La continuité subjective générique se maintient à travers n'importe quelle discontinuité objective dans l'existence des êtres conscients.

On pourra consulter avec intérêt le site de Wayne Stewart[308] pour son développement d'une hypothèse très similaire reposant sur des arguments très semblables.

Malgré des avantages indéniables liés à sa simplicité, la thèse phénoménaliste que nous avons présenté dans ce sous-chapitre, paraît peu croyable, même avec beaucoup d'efforts. Le défi qu'elle nous pose : nier notre existence réelle en tant que personnes individuelles. Cela est comparable à nier l'existence du monde qui nous entoure… Malgré sa très bonne adéquation à une vision naturaliste du monde, il faut donc souligner le caractère assez contre-intuitif de ce modèle, rendant difficilement compte de l'identité personnelle, surtout d'un point de vue subjectif. C'est surtout problématique au sein de notre société occidentale, qui prise les valeurs de l'ego. Néanmoins, cognitivement, cette théorie est à même de rendre compte de notre *impression* d'identité personnelle, par les phénomènes de clôture phénoménologique qui nous empêche d'expérimenter ce que cela fait d'être quelqu'un d'autre (parce que nous en sommes physiquement séparé), et le sentiment d'avoir toujours été présent, associé à notre mémoire autobiographique, qui nous donnent une *impression* de stabilité au cours du temps.

### *B. Deuxième modèle : théorie réaliste ontologique de l'identité personnelle.*

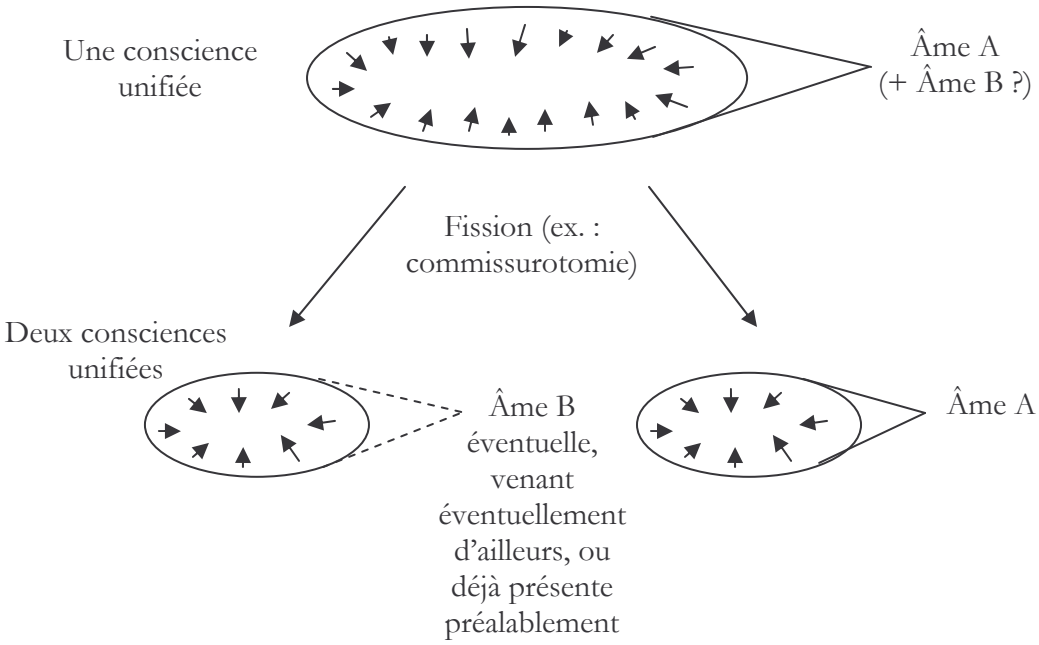
On peut enraciner ce deuxième modèle dans le premier, dont il cherche à corriger certains défauts, par un recours à certains présupposés dualistes hérités de Platon, Descartes, Kant, Eccles, etc. mais surtout de la pensée hindouiste (comme nous allons le

[308] STEWART, Wayne, *Metaphysics by default* [en ligne] : http://mbdefault.org/

voir, l'*atman* hindoue impersonnelle, plus que l'âme chrétienne personnelle, semble correspondre à ce modèle).

Partant du constat que la majorité des impressions ne sont pas des illusions ou des hallucinations, ou du moins qu'elles ne sont pas considérées comme telles et qu'il serait contre-intuitif de soutenir le contraire, ce modèle suppose qu'il existe « quelque chose » en plus (que nous pouvons appeler une *âme*), comme soubassement réel de nos impressions d'identité personnelle, dans la majorité des cas. Ainsi, pendant que le modèle phénoménaliste de l'identité personnelle rend compte parfaitement de notre impression de l'identité personnelle, le modèle réaliste ontologique le complète en posant une âme immatérielle qui rende compte, réellement, sur un plan ontologique, non plus, cette fois-ci, de notre impression de l'identité personnelle, mais de l'identité personnelle en tant que telle, dont notre impression devient une perception véridique. Se faisant, il évite de faire de nous des êtres continuellement en proie aux illusions et hallucinations.

On peut représenter ce modèle schématiquement de la sorte :




Il s'agit d'une forme de spiritualisme, mais minimaliste, car l'âme postulée ne serait pas le siège de notre personnalité d'être humain qui s'appelle Pierre, Paul ou Jacques, mais seulement le « soi témoin », le « présent spécieux », que nous avons définis précédemment. Ses propriétés se réduiraient à :

- Unicité

- Identité trans-temporelle

- Indivisibilité

- Indétermination de contenu

- Liberté

- Conscience phénoménale

- Eternité

Il y aurait un sens à se demander ce qu'il y avait avant la naissance ou ce qu'il y aura après la mort, mais ce sens serait purement subjectif.

Si cette thèse rend bien compte des aspects subjectifs de l'identité personnelle, on peut, cependant, se demander si elle ne fait pas au prix fort : celui de l'ajout, au mobilier ontologique du monde, d'entités inobservables et peu intelligibles. De plus, elle ne donne pas de critères pratiques pour l'identification et la réidentification des personnes. Du point de vue, épistémologique, du point de vue de l'identité objective, ce modèle est très faible. C'est pourquoi on peut préférer le modèle suivant.

### *C. Troisième modèle : théorie physicaliste intégrative de l'identité personnelle.*

Ce modèle s'inspire de la plupart des auteurs contemporains et modernes et tente de les unifier en une représentation riche et intégrative de l'identité personnelle. Les facteurs invoqués sont divers mais sont hiérarchisés. Tous n'ont pas la même importance dans la constitution du « moi ».

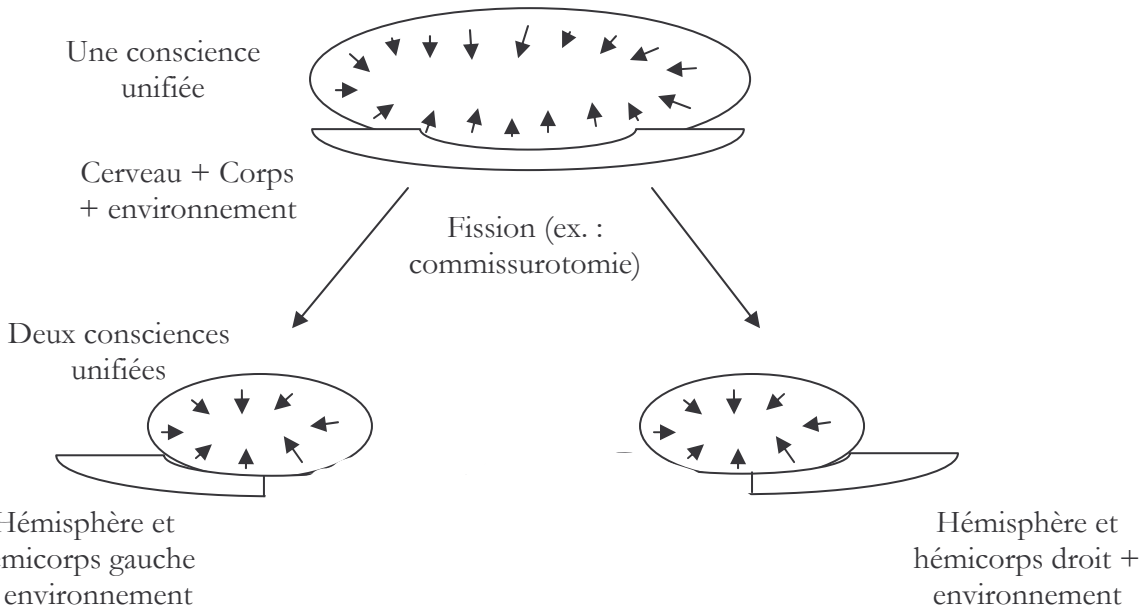
Selon cette théorie, l'identité personnelle dépend des paramètres suivants (les numéros entre parenthèse, à droite, donnent l'ordre d'importance du critère, dans la constitution du *moi*) :

- Critères empiriques :
    - Critères internes :
        - Corps et cerveau (1)
        - Conscience de soi en première personne et capacité de faire des choix selon des préférences (4)
    - Critères externes :
        - Environnement physique (1)
        - Environnement social (3)
        - Communauté linguistique et culturelle (4).
- Continuité et connexité temporelles (appuyées par une ontologie perdurantiste). (2)

On peut justifier le poids relatif donné à ces différents critères de la manière suivante : pour qu'il y ait une personne X ou Y, il faut d'abord un corps muni d'un cerveau

(1) – on voit mal ce que serait une conscience de soi abstraite de tout enracinement matériel – et ce corps est nécessairement situé en un endroit de l'espace à un temps donné, ce qui définit, relativement à lui, un environnement qui, sous sa plus simple expression, est au moins physique (1). Ensuite, au cours du temps, ce corps peut changer de configuration et de localisation, ce qui définit une connexité et une continuité temporelles (2). La présence d'autres individus avec lesquels interagir (3), initie le co-développement, d'une part, d'une communauté culturelle, voire linguistique (5), d'autre part, d'une conscience de soi en première personne, d'une faculté de choix motivé (5).

On peut représenter ce modèle schématiquement de la sorte :




Pour ce modèle, l'identité personnelle se construit progressivement, en parallèle du développement embryonnaire, fœtal, infantile, adolescent. Puis elle se disloque progressivement avec divers maladies et accidents, avec la vieillesse, pour se dissoudre

finalement à la mort. Toutefois, il est peut-être abusif de parler de pur néant avant et après la vie, car si on pousse jusqu'au bout ce modèle, on en vient à prendre conscience du flux et reflux permanent qu'est la vie. Les continuités et connexités temporelles, qu'on les considère des points de vue de l'organisme, de la conscience, de la culture ou de l'environnement physique et social, commencent bien avant le berceau et continuent par-delà le tombeau. Il n'y a pas de frontière nette et étanche qui viendrait séparer un « avant » d'un « après », un « nous-même » d'un « quelqu'un d'autre ». La singularité de notre moi doit se comprendre tel que la comprenait le philosophe indien Krishnamurti. Dans *Krishnamurti ou la révolution du réel*, René Fouère[309] nous expose cette conception :

> « Se dressant contre la conviction commune, le sentiment commun, que l'individu possède une existence séparée, Krishnamurti ne cesse de répéter que cet individu n'a pas de réalité hors des relations qu'il soutient avec le milieu ou avec ses propres éléments constitutifs. "Etre, dit-il, c'est exister dans des rapports." S'éloignant, du moins en apparence, des perspectives du cogito cartésien, il proclame obstinément: "Vous existez parce que vous êtes relié." Certes, dans sa notion aiguë, le moi est insurmontable persuasion d'existence séparée, mais cette persuasion n'est pas entièrement fondée. Elle implique méprise et malentendu. L'individualité est bien une solitude spirituelle ou psychologique, mais elle ne cesse pas pour autant de s'enraciner, objectivement, dans l'univers des choses, des pensées et des oeuvres. Si l'individu est unique, c'est seulement en ce sens que, de tous les sucs nourriciers du monde, il tire un miel inimitable. En d'autres termes, il n'est que l'originale synthèse de ses rapports avec le monde, la conscience singulière qu'il en prend. Sa vision et sa perspective lui sont propres, mais elles n'ont pas d'existence intrinsèque: elles n'existent qu'en fonction d'un monde et en association permanente avec lui. Ce qui définit l'individu, c'est le caractère unique de son expérience, mais non pas sa séparation - en quelque sorte "matérielle" - d'avec autrui, sur quelque plan et à quelque niveau que ce soit. En résumé, l'unicité – Krishnamurti s'est servi spécifiquement de ce

[309] FOUERE, René. *Krishnamurti ou la révolution du réel.*

> terme – est réelle, la séparation illusoire : "Nous sommes différents, mais non séparés" me disait un jour Krishnamurti. »

Ainsi, l'on peut envisager plusieurs manières dont notre identité – dans ce modèle – pourrait se poursuivre après notre décès :

- **Par notre descendance :** rigoureusement parlant, il y a une parfaite continuité corporelle dans le processus par lequel un gamète – lequel contient la moitié de notre génome – en vient à former un zygote, puis un embryon, un fœtus, un enfant, par fusion avec un autre gamète du sexe opposé[310].

- **Par notre œuvre :** dans une perspective externaliste, les traces que nous laissons de notre passage peuvent être considérées comme des « bouts » de nous-même qui nous survivent. Il ne s'agit pas d'affirmer, claironnant, qu'Emmanuel Kant survit littéralement dans la *Critique de la Raison Pure*, mais l'influence que ses ouvrages peuvent avoir sur les consciences font revivre en nous – et cette fois-ci, l'on peut penser que ce n'est pas seulement métaphorique – une partie de lui-même. Car il y a une continuité psychologique, et surtout idéologique, qui se met en place par cette intermédiaire. Les idées de Kant, à présent séparées de leur

[310] Nous avons vu, cependant, qu'une continuité corporelle pouvait ne pas suffire à préserver l'identité personnelle, si elle n'est pas d'un certain type (que nous avons qualifié de « continuité corporelle régénératrice » et de « continuité corporelle sub-personnalisante »). Précisément, dans le processus de la procréation, il semble que l'identité numérique ne soit pas préservée : nous ne sommes pas nos enfants, puisqu'en règle générale, nous coexistons avec eux un certain temps. Toutefois, nous pouvons le devenir ou devenir nos petits-enfants, arrières petits-enfants, etc., après notre mort. L'idée a de quoi surprendre, mais ce que nous voulons dire n'est rien de plus que ce que défend Thomas W. Clark. A notre mort, notre flux tumultueux de subjectivité se « déverse », en celui de nos descendants, et il y poursuit son existence, quoique radicalement transformé, à la manière dont un fleuve se poursuit, méconnaissable, dans la mer, puisque c'est l'eau qui était dans le fleuve se retrouve à présent dans la mer.

source originelle par plusieurs siècles, poursuivent leur existence, de personne en personne, par l'intermédiaire des livres et des paroles (et à présent, aussi grâce à l'Internet). C'est la théorie des *mèmes* de Richard Dawkins[311], Susan Blackmore[312], Daniel Dennett[313], etc. qui sont des gens que l'on ne peut guère suspecter pourtant de complaisance avec le spiritualisme ! Selon cette théorie, exposée pour la première fois dans *Le Gène Egoïste*, par Richard Dawkins[314], un mème est une unité réplicative d'information culturelle présente dans un cerveau, et échangeable au sein d'une société. Les mèmes seraient responsables de l'évolution culturelle et de celle de certains comportements acquis, par un processus de variabilité et de sélection naturelle (ou plus exactement culturelle), au même titre que les gènes sont responsables de l'évolution biologique. La « survie » d'un mème est bâtie sur leur capacité à provoquer un comportement imitable. La science des mèmes est la mémétique. Tout le débat autour des mèmes consiste principalement à juger de la pertinence de cette analogie avec les gènes et de son pouvoir explicatif. En tout état de cause, la notion de mème ne contredit pas une vision naturaliste du monde, s'y accorde même plutôt bien, et est assez « séduisante ». De plus, sans vouloir trancher le débat, il faut reconnaître que quelques bons arguments sont à son appui[315].

311 DAWKINS, Richard, *Le Gène Egoïste.*

312 BLACKMORE, Susan. *The Meme Machine.*

313 DENNETT, Daniel Clement. *Darwin's Dangerous Idea.*

314 DAWKINS, Richard, *Le Gène Egoïste.*

315 Voir BLACKMORE, Susan. *The Meme Machine.*

- **Dans la mémoire collective :** Une autre idée, fortement liée à la mémétique, s'enracine également dans l'externalisme actif[316] et la thèse du super-organisme développée, entre autres, par Joël de Rosnay[317] et les membres du collectif *Principia Cybernetica*[318]. C'est l'idée selon laquelle les groupes humains (de deux personnes ou plus, jusqu'à l'ensemble de la planète) constituent de vastes réseaux de communication (surtout depuis l'émergence de l'Internet) comparables aux réseaux de neurones du cerveau. Analogiquement, on serait fondé à imaginer une conscience collective, qui serait le pendant, à un niveau social, de la conscience individuelle au niveau psychobiologique. Il serait concevable d'imaginer survivre à travers cette conscience globale, lorsque notre conscience individuelle s'éteindrait, dans la mesure où notre présence aurait imprégné cette conscience collective, qui se trouverait dès lors en continuité psychologique avec elle. Autrement dit, l'on survivrait littéralement à travers le souvenir de nos proches et par extension de la société entière, si on l'a suffisamment marquée. Pour prendre une image approximative que nous espérerons éclairante : on peut dire qu'après que l'on ait amputé un bras à quelqu'un, ce bras survit malgré tout dans la conscience de cette personne, dans la mesure où subsiste en elle l'impression d'avoir toujours ce bras manquant (membre fantôme). Cette impression résulte du fait que le cerveau se « souvient » du bras. Celui-ci l'a marqué dans sa structure neuronale, de par l'activité qu'il déployait et la sensibilité dont il était le siège. On dit que le cerveau fait preuve de plasticité. Cette même plasticité tend

316 Voir notre partie sur l'externalisme actif

317 DE ROSNAY, Joël, *Le Macroscope.*

318 HEYLIGHEN, Francis, JOSLYN, Cliff et TURCHIN, Valentin, *Principia Cybernetica Web* [en ligne] : < http://pespmc1.vub.ac.be/ >

> subséquemment à réduire la trace neuronale, donc le « souvenir » cérébral de ce bras, au cours du temps, réduisant par conséquent l'impression de membre fantôme. Toutefois, cela prend du temps. La force de l'habitude induit en effet une certaine inertie des représentations neuronales et donc mentales. Il est parfois aussi difficile, sinon plus, de désapprendre que d'apprendre…

On le voit, ce modèle rend bien compte des aspects, tant objectifs que subjectifs de l'identité personnelle, et cela, avec parcimonie ontologique. Il tient compte de la richesse des interactions qui se jouent au sein de l'individu et avec son environnement. Il ne cherche pas à simplifier outrancièrement les choses parce qu'une réponse simple serait plus plaisante.

On peut se demander néanmoins si les critères évoqués ne sont pas davantage pertinents d'un point de vue épistémologique qu'ontologique. Lorsque l'on a cité tous ces critères, a-t-on véritablement épuisé l'essence d'une personne ? Pour nous, est-ce vraiment ce genre de choses comme le corps, la localisation spatiale, la continuité temporelle, etc. qui importent ? Bien sûr, le modèle fait aussi entrer en ligne de compte les souvenirs, la conscience de soi en première personne, la culture, la personnalité, etc. Mais il ne leur donne pas la première place.

D'un autre côté, il faut reconnaître que l'âme impersonnelle du modèle ontologique réaliste n'est guère mieux placée, de par son dépouillement de propriétés. On aimerait y inclure d'autres choses, comme la conscience de soi, la mémoire, les émotions, etc. Certes, elle est libre et identique à elle-même tout au long de son existence. Elle satisfait donc les exigences les plus impérieuses, elle synthétise la quintessence de la personne. Mais, même si

nous nous identifions réellement à elle, nous *aimons*, en tant qu'êtres humains, nous identifier à plein d'autres choses, futiles d'un point de vue métaphysique, tel que notre carrière, notre intelligence, notre beauté, notre prestige…

### *D. Synthèse.*

Nous avons présenté ces modèles comme trois options distinctes, laissant entendre qu'elles étaient peut-être exclusives l'une de l'autre, et qu'il nous fallait choisir. Néanmoins, point n'est besoin de choisir entre ces trois modèles, de même que l'on n'a pas forcément à choisir entre Dieu et l'évolution. La théorie phénoménaliste rend bien compte de notre sentiment de nous-même et des autres, la théorie réaliste ontologique fonde ontologiquement notre identité, donnant ainsi un contenu substantiel à nos intuitions, quant à la théorie physicaliste intégrative, elle fournit des critères, épistémologiquement utiles, à l'identification et à la réidentification des personnes.

L'identité personnelle, au mieux, doit, au pire, peut, se comprendre à trois niveaux autonomes, différents, mais complémentaires, qui s'emboîtent selon la progression suivante, d'un point de vue épistémologique :

1) **Un niveau physicaliste intégratif :** Ce niveau permet de connaître l'identité personnelle objectivement, de manière vérifiable. Bien que la précision ne soit pas de 100%, il s'agit du niveau qui permet néanmoins d'aller le plus loin au fond des choses. Comme le souligne Jean Bricmont[319], *« l'approche scientifique (en y incluant la connaissance ordinaire) nous donne les seules*

[319] BRICMONT, Jean. « Science et religion : l'irréductible antagonisme »

*connaissances objectives auxquelles l'être humain a réellement accès. Si l'approche scientifique nous donne une vision partielle de la réalité, c'est parce que nous n'avons pas accès, de par notre nature finie, à la réalité ultime des choses. […] [I]l y a une grande différence entre dire que la science nous donne une description complète de la réalité et dire qu'elle en donne la seule connaissance accessible à l'être humain ».* Ensuite on a :

2) **Un niveau phénoménaliste :** De ce niveau, tout ce que l'on peut connaître est, non pas notre identité (ce que nous sommes), mais notre *différence* (ce que nous ne sommes pas). Nous ne sommes pas cette personne là-bas en train de faire ceci ou cela, puisque ce n'est pas sa bouche qui bouge lorsque nous disons cela et que ce n'est pas non plus de sa bouche que sortent ces paroles. Mais qui sommes-nous ? La phénoménologie ne peut pas nous le dire. Enfin, le niveau le plus faible, épistémiquement est :

3) **Un niveau réaliste ontologique :** De ce niveau, on ne peut rien dire du tout.

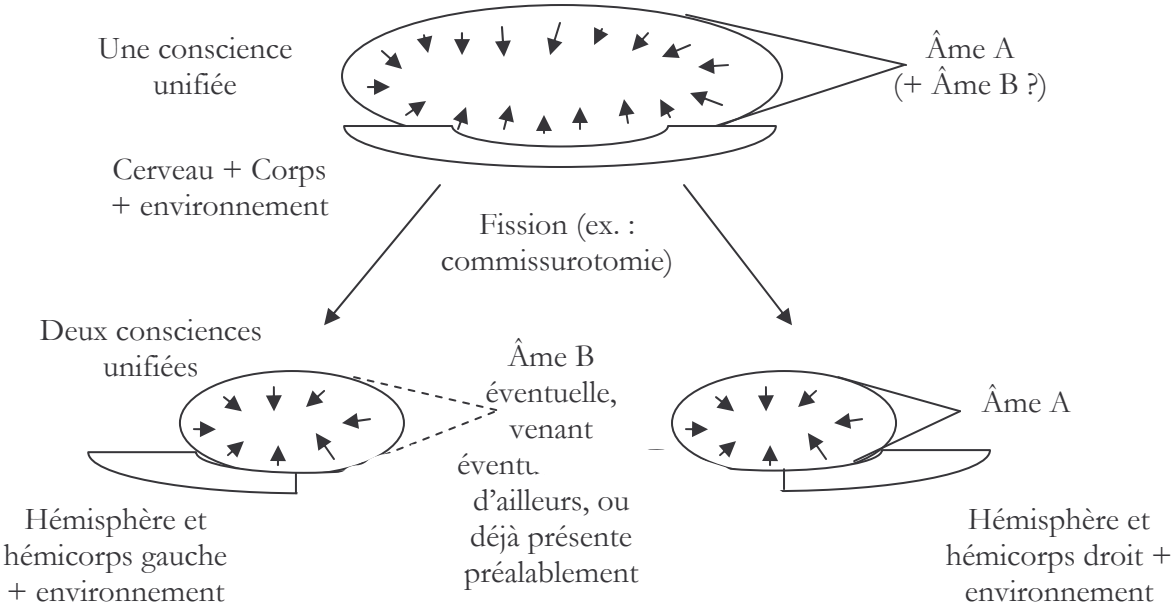
Toutefois, sur un plan ontologique, nous pouvons proposer un ordre inversé. Le primat ontologique – ce qui définirait réellement ce que nous sommes – reviendrait à :

1) **Un niveau réaliste ontologique :** Lequel engloberait :

2) **Un niveau phénoménologique :** Qui serait ontologiquement second, et qui engloberait lui-même :

3) **Un niveau physicaliste intégratif :** Qui serait ontologiquement troisième et dernier.

On peut représenter schématiquement cette synthèse de la sorte :




Nous avons commencé par critiquer la notion de néant, montrant son incohérence, tant du point de vue logique que sémantique. Nous avons ensuite apporté quelques arguments plus positifs, en faveur de la survie du sujet, et même de son éternité. Dans ce troisième chapitre des annexes, le lien entre le problème de la survie et celui de l'identité personnelle s'est fait plus clair. Ces deux questionnements sont complémentaires et s'éclairent réciproquement.

Non seulement la thèse de la survie n'est incompatible avec aucun modèle et est appuyée par certains, mais encore, aucun de ces modèles n'est exclusif. Il n'y a pas à choisir. Parce que, tout simplement, ils n'opèrent pas au même niveau.

On le voit également, la critique de la notion de néant que nous avons esquissée, fournit un appui à la notion de « soi » témoin, laquelle est fondamentale à une compréhension métaphysique du problème de l'identité personnelle, seule à même de justifier un recours aux expériences de pensée type « science fiction », et à donner un sens à toutes les interrogations contrefactuelles relatives à l'identité personnelle à travers les mondes possibles.

# Bibliographie et index

« Si j'écrivais un livre intitulé *Le monde tel que je l'ai trouvé*, je devrais y faire aussi un rapport sur mon corps, et dire quels membres sont soumis à ma volonté, quels n'y sont pas soumis, etc. Ce qui est en effet une méthode pour isoler le sujet, ou plutôt pour montrer que, en un sens important, il n'y a pas de sujet : car c'est de lui seulement qu'il ne pourrait être question dans ce livre. »

(Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, § 5.631)

# _BIBLIOGRAPHIE_

## I) Références des citations utilisées en épitaphes :

MORRISSEY (2004). “My Life is a Series of People Saying Goodbye”. *In:* You Are the Quarry [Deluxe Edition]. Attack/Sanctuary. Album musical.

GOUNELLE, André (1981). « Le dynamisme créateur de Dieu. Essai sur la théologie du Process ». *Etudes théologiques et religieuses*, 56ème année, Hors-Série. 192 p.

LODGE, David (2001). Pensées secrètes, traduit de l’anglais par MAYOUX, Suzanne V. Ed. Rivages, Paris, 2002. 402 p.

CARROLL, Lewis (1865). *Les Aventures d’Alice au Pays des Merveilles*. Ed. J’ai Lu, Paris, 2004. 91 p.

## II) Ouvrages papier :

### A) Ouvrages généraux :

*A Companion to Metaphysics* (1995). Edité par GUTTENPLAN, Samuel, édition Blackwell Publishing 2004, collection Blackwell Companions to Philosophy, 540 p.

*A Companion to the Philosophy of Mind* (1994). Edité par KIM, Jaegwon et SOSA, Ernest, édition Blackwell Publishing 2004, collection Blackwell Companions to Philosophy, 642 p.

*La Sainte Bible*, traduite par SEGOND, Louis. Trinitarian Bible Society (The Chaucer Press), London.

KINOT Françoise (2000 à 2002). Collection Philosophie, Philosophie contemporaine n°XI, Philosophie des lumières n°VI et VIII. Editions France Loisirs, Paris, 12 volumes.

**B) Monographies (Roman, traités, ouvrages critiques et essais philosophiques, script) :**

1) Roman :

ABBOTT, Edwin Abbott (1884). *Flatland: A Romance of Many Dimensions*, Dover Thrift Editions, 1992, 96 p.

2) Traités, ouvrages critiques et essais philosophiques :

AUGUSTIN, Saint (401). *Confessions*. Seuil, Paris, 2004, 405 p.

AYERS, Michael J. (1990). *Locke*. Routledge, London. 98 p.

BERKELEY, Georges (1713). *Trois Dialogues Entre Hylas Et Philonous*. Flammarion, Paris, 1998. 297 p.

BLACKMORE, Susan Jane (2000). *The Meme Machine*. Oxford University Press, Oxford. 288 p.

BOECE, *Courts traités de théologie*, traduit de l'anglais par MERLE, Hélène de Boethius, *The Theological Tractates*, éd. G.P. Goold, Cambridge (Mass.), Harvard University Press, The Loeb Classical Library, 1973. Cerf, Paris, 1991. 154 p.

BUTLER, John (1736). *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature*. Kessinger Publishing, Whitefish, 2003, 344 p.

CARNAP, Rudolf (1928). *La construction logique du monde*, introduction de SCHWARTZ, Elisabeth ; traduit de l'allemand par RIVAIN, Thierry de *Der logische Aufbau der Welt* (1928) ; révision par SCHWARTZ, Elisabeth, Vrin, Paris, 2002, 370 p.

CHANGEUX, Jean Pierre (1983). *L'Homme neuronal*. Fayard, Le temps des sciences, Paris. 419 p.

CHAUVIER, Stéphane (2003). *Qu'est-ce qu'une personne ?* Vrin, Paris. 124 p.

CHOMSKY, Noam (2001). *Le Langage et la pensée*. Petite bibliothèque Payot, Paris. 144 p.

DAMASIO, Antonio Rosa (1999). *Le Sentiment même de soi*, traduit de l'anglais par LARSONNEUR, Claire et TIERCELIN, Claudine, Odile Jacob, Paris, 2002, 480 p.

D'AQUIN, Saint Thomas. *Somme théologique*. Paris, Edition de la revue des jeunes, 1953, 473 p.

DARWIN, Charles (1859). *L'Origine des Espèces*, traduit de l'anglais par Barbier, E.,

préface de Drouin, Jean-Marc. Flammarion, Paris, 1999, 608 p.

DAVIDSON, David (1980). *Essays on actions and events*. Oxford University Press, Oxford, 2001, 352 p.

DAWKINS, Richard. (1997). *The Blind Watchmaker*. W. W. Norton, New York.

DAWKINS, Richard. (1989). *Le Gène Egoïste*. Odile Jacob, Paris, 2003, 200 p.

DE ROSNAY, Joël (1975). *Le Macroscope*. Seuil, Paris, 1977, 346 p.

DENNETT, Daniel Clement (1991). *Consciousness Explained*, illustré par WEINER, Paul. Little Brown, Boston, 511 p.

DENNETT, Daniel Clement (1996). *Darwin's Dangerous Idea*. Penguin, London, 592 p.

DESCARTES, René. (1647). *Méditations métaphysiques*. Flammarion, Paris, 1993, 574 p.

ECCLES, John Carew. (1981). *Evolution du cerveau et création de la conscience*. Flammarion, Paris, 368 p.

ECCLES, John Carew (1997). *Comment la conscience contrôle le cerveau*. Fayard, Paris, 256 p.

FERRET, Stéphane (1993). *Le philosophe et son scalpel*. Editions de Minuit, Paris. 130 p.

FERRET, Stéphane (1998). *L'identité*. Flammarion, Paris, 240 p.

FODOR, Jerry A., MILLER, George A. and LANGENDOEN, D. Terence (1975). *The Language of Thought*. Harvard University Press, Cambridge (Massachusetts), 224 p.

HEGEL, Georg Wilhelm Friedrich (1816). *Science de la Logique,* traduit de l'allemand, présenté et annoté par LABARRIERE, Pierre-Jean et JARCZYK, Gwendoline. Aubier, Paris, 1987, 382 p.

HUME, David (1740). *Traité de la nature humaine (2 tomes)*. Traduction, préface et notes de LEROY, André, Aubier, Paris, 1946, 766 p.

HUME, David (1779). *Dialogues sur la religion naturelle*, traduit de l'anglais et commenté par MALHERBE, Michel. Vrin, Paris, 2005, 384 p.

HUSSERL, Edmund (1928). *Leçons pour une phénoménologie de la conscience intime du temps*, 2ème édition, traduit de l'allemand par DUSSORT, Henri. Presses Universitaires de France, Paris, 1983, 205 p.

JAMES, William (1890). *The Principles of Psychology*, with introduction by MILLER,

George A. Harvard University Press, Cambridge (Massachusetts), 1983, 1328 p.

KANT, Emmanuel. (1787). *Critique de la Raison pure*, traduction française avec notes, par TRAMESAYGUES, A. et PACAUD, B., préface de SERRUS C. Presses Universitaires de France, Quadrige, 6ème édition, Paris, 2001, 616 p.

KANT, Emmanuel (1788). *Critique de la raison pratique*, traduit par FUSSLER Jean-Pierre, Flammarion, Paris, 2003, 473 p.

KRIPKE, Saul Aaron (1980). *Naming and Necessity*. Harvard University Press, Cambridge (Massachusetts), 172 p.

LAPLACE, Pierre-Simon (1825). *Essai Philosophique sur les Probabilités*. Ed. Gauthier-Villars, Paris, 1921.

LATOUR, Bruno et WOOLGAR, Steve (1988). *La Vie de laboratoire : la Production des faits scientifiques*. La Découverte, Paris.

LEIBNIZ, Gottfried Wilhelm (1714). *Monadologie*. Edition annotée par BOUTROUX, Emile, Delagrave, Bussigny-près-Lausanne (Suisse). 2002, 231 p.

LEIBNIZ, Gottfried Wilhelm. (1866). *Nouveaux essais sur l'entendement humain*. Chronologie, bibliographie, introduction et notes par BRUNSCHWIG, Jacques. GF Flammarion, Paris, 1990, 442 p.

LOCKE, John (1694). *Essai sur l'entendement humain*, traduit de l'anglais, préfacé, annoté et indexé, par VIENNE, Jean-Michel. Vrin, Paris, 2001, 640 p.

LOVELOCK, James Ephraim (1979). *Gaia: A New Look at Life on Earth*. Oxford University Press, New York, 1987, 176 p.

LUCRECE, *De la Nature des choses*. Sous la direction de BALAUDE, Jean-François. Le Livre de Poche, Paris, 2002, 695 p.

MACH, Ernst (1883). *The Science of Mechanics*, translated by MCCORMACK, Thomas J., sixth edition, Open Court, LaSalle Illinois, 1960, 668 p.

MAHADEVAN, T. M. P. (1982). *Qui suis-je ?*, traduit de l'anglais par BRAITENBERG, Eléonore. Pondicherry, All India Press, 2006.

MARGULIS Lynn and SAGAN, Dorion (1986). *Microcosmos: Four Billions Years of Evolution from Our Microbial Ancestors*. Summit Books, New York, 301 p.

MILL, John Stuart. (1861). *L'Utilitarisme*, Flammarion, Paris, 1999, 181 p.

MONOD, Jacques (1971). *Le hasard et la nécessité*. Seuil, Paris, 1973, 244 p.

NOZICK, Robert (1981). *Philosophical Explanations*. Harvard University Press,

Cambridge, 784 p.

NOZICK, Robert (1993). *The Nature of Rationality.* Princeton University Press, Princeton, 1994, 242 p.

OLSON, Eric (1997). *The Human Animal: Personal Identity Without Psychology.* Oxford University Press, Oxford, 200 p.

PAINE, Thomas (1796). *The Age of Reason*. Carol Publishing Corporation, New York, 1976.

PARFIT, Derek (1984). *Reasons and Persons*. Oxford University Press, Oxford, 1986, 543 p.

PASCAL, Blaise (1670). *Les Pensées : Classées selon les indications manuscrites de PASCAL*, préfacées et annotées par KAPLAN, Francis. Editions du Cerf, Paris, 2005, 705 p.

PEARS, David (1986). *La pensée-Wittgenstein : du Tractatus aux Recherches Philosophiques*, Aubier, Paris, 1993, 536 p., traduit de l'anglais : *The False Prison, vol. II*, Clarendon Press, Oxford, par CHAUVIRÉ, Christiane.

PLANTINGA, Alvin (1967). *God and Other Minds: A Study of the Rational Justification of Belief in God*. Cornell University Press.

POPPER, Karl Raimond et ECCLES, John Carew (1977). *The Self and Its Brain*. Springer-Verlag, Berlin-Heidelberg, 597 p.

POUIVET, Roger (1997). *Après Wittgenstein, saint Thomas*. Presses Universitaires de France, Paris. 128 p.

QUINE, Willard Van Orman (1993). *La poursuite de la vérité*. Seuil, Paris, 1999, 153 p.

QUINE, Willard Van Orman (1980). *From a Logical Point of View: Nine Logico-Philosophical Essays*, Second Revised Edition. Harvard University Press, Cambridge (Massachusetts), 2003, 200 p.

RAMACHANDRAN, Vilayanur S. et Blakeslee, Sandra (2000). *Le Fantôme Intérieur*, traduit de l'anglais par GARENE, Michèle et préfacé par SACKS, Olivier. Odile Jacob, Paris, 2002, 368 p.

REID, Thomas (1785). *Essais sur les facultés de l'esprit humain (4 volumes)*, traduit de l'anglais et publié par JOUFFROY, Théodore Simon, 1829, 1761 p.

REID, Thomas (1785). *Essais sur les pouvoirs intellectuels de l'homme*, Edition Hamilton, 1845.

RUSSELL, Bertrand. (1912). *Problèmes de philosophie*, traduit de l'anglais par RIVENC,

François. Paris, Payot, 1989, 190 p.

RYLE, Gilbert (1949) *The Concept of Mind*. University Of Chicago Press: New Univer edition, Chicago, 2000, 348 p.

SEARLE, John Richard (1995). *La redécouverte de l'esprit*, traduction de l'anglais par TIERCELIN, Claudine. Collection NRF Essais Gallimard, Paris, 353 p.

SHOEMAKER, Sidney (1996). *The First-Person Perspective and Other Essays*. Cambridge University Press, Cambridge, 296 p.

SOKAL, Alan et BRICMONT, Jean (2005). *Pseudosciences et postmodernisme. Adversaires ou compagnons de route ?* Odile Jacob, Paris, 224 p.

SPINOZA, Baruch (1677). *L'Ethique*, Gallimard, Paris, 1994, 398 p.

STRAWSON, Peter F. (1990). *Individuals. An Essay in Descriptive Metaphysics*. Routledge, London, 256 p.

SWINBURNE, Richard. (1996). *Is there a God?* Oxford University Press, Oxford, 160 p.

TEILHARD DE CHARDIN, Pierre. (1955). *Le phénomène humain*. Seuil, Paris, 1970, 320 p.

VAN INWAGEN, Peter (1983) *An Essay on Free Will*. Oxford University Press, USA, 1986, 254 p.

WEARING, Deborah (2006). *Forever Today*. Corgi, 400 p.

WIGGINS, David (1980). *Sameness and Substance*. Cambridge University Press, Cambridge (Massachusetts), 2001, 274 p.

WILKES, Kathleen V. (1988). *Real People: Personal Identity without Thought Experiments*. Clarendon Press, Oxford, 1994, 264 p.

WITTGENSTEIN, Ludwig (1922). *Tractatus logico-philosophicus*, traduit de l'allemand par GRANGER Gilles-Gaston, notes de GRANGER Gilles-Gaston, introduction par RUSSELL, Bertrand. Gallimard, Paris, 1993, 122 p.

WITTGENSTEIN Ludwig (1953). *Recherches philosophiques*, traduit de l'allemand par DASTUR, Françoise, ELIE, Maurice, GAUTERO, Jean-Luc, JANICAUD, Dominique, RIGAL, Elisabeth, de *Philosophische Untersuchungen*, avant-propos et apparat critique de RIGAL, Elisabeth. Gallimard, Paris, 2004, 368 p.

WITTGENSTEIN, Ludwig (1958). *Le Cahier bleu et le Cahier brun*, traduit de l'anglais par GOLDBERG, Marc et SACKUR, Jérôme, préfacé par IMBERT, Claude. Gallimard, Paris, 1996, 314 p.

WITTGENSTEIN, Ludwig. (1967). *Zettel*, edited by ANSCOMBE, Gertrude Elisabeth Margaret and VON WRIGHT, Georg Henrik, translated by ANSCOMBE, Gertrude Elizabeth Margaret, Basil Blackwell, Oxford.

WITTGENSTEIN, Ludwig (1969). *De la certitude*. Gallimard, Paris, 1976, 152 p.

WITTGENSTEIN, Ludwig (1980). *Remarques sur la philosophie de la psychologie I*, traduit de l'allemand par GRANEL, Gérard. Trans-Europ-Repress Editions, Mauzevin (France), 1989.

3) Lettres philosophiques :

EPICURE, Lettre à Ménecée.

4) script cinématographique :

HANDKE Peter et WENDERS Wim (1989) traduit de l'allemand par PETIT, Dominique et EISENSCHITZ, Bernard. *Les Ailes du désir*, éditions Jade-Flammarion, Le Chesnay, 174 pages

**C) Articles de périodiques (recueil d'articles, revue, magazine, etc) :**

ASPECT, Alain, DALIBARD, Jean et ROGER, Gérard (1982). "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers". *Physical Review Letters*, 49(25) : 1804-1807.

ASPECT, Alain, GRANGIER, Philippe et ROGER, Gérard (1982). "Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities". *Physical Review Letters*, 49(2) : 91-94.

BROWN, James Robert (2004). "Peeking into Plato's Heaven". *Philosophy of Science*, 71:1126-1138.

BURGE, Tyler (1979). "Individualism and the Mental". *In:* FRENCH, Peter A., UEHLING, Thomas E. and WETTSTEIN, Howard K. (Eds.). *Midwest Studies in Philosophy IV: Studies in Metaphysics*. University of Minnesota Press, Minneapolis, 73-121.

COHNITZ, David (2003). "Personal Identity and the Methodology of Imaginary Cases". *In:* PETRUS, Klaus (Ed.) (2003). *On Human Persons*. Ontos Verlag, p. 145-181.

DENNETT, Daniel Clement (1987) « Où suis-je ? » *In :* HOFSTADTER, Douglas et DENNETT, Daniel Clement (Eds.). *Vues de l'esprit : fantaisies et réflexions sur l'être et l'âme*, traduit de l'anglais par HENRY, Jacqueline. InterEditions, Paris, p. 219-230.

DENNETT, Daniel Clement and KINSBOURNE, Marcel (1992). "Time and the Observer : The Where and When of Consciousness in the Brain". *Behavioral and Brain Sciences*, 15 : 183-247.

DEPAUL, Michael R. and RAMSEY, William (1998). *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry*. The British Journal for the Philosophy of Science, 2001, 52:151-158.

FRIED, I., WILSON, C. L., MACDONALD, K. A. and BEHNKE, E. J. (1998). Electric current stimulates laughter. *Nature*, 12;391(6668):650.

HARDING, Douglas (1987). « Vivre sans Tête ». *In :* HOFSTADTER, Douglas et DENNETT Daniel Clement (Eds.). *Vues de l'esprit : fantaisies et réflexions sur l'être et l'âme*, traduit de l'anglais par HENRY, Jacqueline. InterEditions, Paris, p. 32-38.

KUHN, Thomas. (1964). "A Function for Thought Experiments", *In:* KUHN, Thomas (1977). *The Essential Tension*, University of Chicago Press, Chicago, p. 240-265

LEWIS, David Kellogg (1976). "Survival and Identity", *In:* RORTY, Amélie Oksenberg (Ed.). *The Identities of Persons*. Berkeley, California, p. 17-40.

MARCEL, Anthony J. (1993). "Slippage in the Unity of Consciousness". *In:* BOCK, Gregory R. and MARSH, Joan (Eds.). *Experimental and Theoretical Studies of Consciousness.* John Wiley et Sons, Chichester, p. 168-179.

MARCEL, Anthony J. (1994). "What is Relevant to the Unity of Consciousness?" *In* : PEACOKE Christopher (Ed.), *Objectivity, Simulation and the Unity of Consciousness*. Oxford University Press, Oxford, p. 79-88.

MARCUS, Eric. (2004). "Why Zombies are Inconceivable". *Australasian Journal of Philosophy*, 82:477-490.

MCTAGGART, John McTaggart Ellis (1993). "The Unreality of Time". *In:* LE POIDEVIN Robin et MCBEATH Murray (Eds.). *The Philosophy of Time*. Oxford University Press, Oxford, 1993:23-24.

MILLIKAN, Ruth Garrett (1993). "Content and Vehicle". *In:* EILAN, Naomi., MCCARTHY, Rosaleen and BREWER, Bill (Eds.). *Spatial Representation.* Blackwell, Oxford, p. 256-268.

NAGEL, Thomas (1974) "What is it like to be a Bat?" *Philosophical Review*, 83 : 435-456.

NAGEL, Thomas (1997) "Brain Bisection and the Unity of Consciousness". *In* : PERRY, John (Ed.). *Personal Identity*. University of California Press, London, 227-245.

PARFIT, Derek (1971). "Personal Identity", *The Philosophical Review*, vol. 80, no. 1, January 1971

PERRY, John (1972). "Can the Self Divide?" *Journal of Philosophy*, 69:463-488.

PERRY, John. (1979). "The Problem of the Essential Indexical". *Nous*, 13:3-21.

POUIVET, Roger (2002). « Epistémologie de la croyance religieuse ». *In :* BOURGEOIS-GIRONDE, Sacha, GNASSOUNOU, Bruno et POUIVET, Roger (Eds.). *Analyse et théologie : Croyances religieuses et rationalité*. Vrin, Paris. 296 p.

PUTNAM, Hilary (1975). "The meaning of 'meaning'". *In*: PUTNAM, Hilary. (1979) *Mind, Language, and Reality*. Cambridge University Press, London, p. 215-271.

QUINE, Willard Van Orman (1972). "Review", *Journal of Philosophy,* 69:491-497.

RUSSELL, Bertrand (1915). "On the Experience of Time". *Monist*, 25 : 212-233.

SCHLICK, Moritz (1913). "Is there Intuitive Knowledge?", *In:* MULDER, Henk L. et VAN DE VELDE-SCHICK, Barbara F.B., *Moritz Schlick: Philosophical Papers, Vol. 1*, D. Riedel, Dordrecht, 1979, p. 141-152.

SCOVILLE, William Beecher and MILNER, Brenda (1957). "Loss of recent memory after bilateral hippocampal lesions". *Journal of Neurology, Neurosurgery and Psychiatry*, 20 : 11-21.

SHOEMAKER, Sidney (1975). "Functionalism and Qualia". *Philosophical Studies*, 27:291-315.

DE SOUSA, Ronald (1976). *Rational Homunculi*. *In*: Rorty A.O., The Identity of Persons, University of California Press, Berkeley, p. 217-238.

SPEMANN, Hans und MANGOLD, Hilde (1924) „Über Induktion von Embryonalanlagendurch Implantation artfremder Organisatoren". *Wilhelm Roux Arch. Entw. Mech. Org.*, 100:599-638.

THOMSON, Judith Jarvis. (1997). "People and Their Bodies", *In*: DANCY, Jonathan (Ed.). *Reading Parfit*. Basil Blackwell, Oxford, p. 202-229.

TREVARTHEN, Colwyn (1974) "Analysis of Cerebral Activities that Generate and Regulated Consciousness in Commissurotomy Patients". *In:* DIAMOND, Stuart J. and BEAUMONT, J. Graham, *Hemisphere Function in the Human Brain*, *Elek Science*, London, 235-261.

TREVARTHEN, Colwyn (1978). “Modes of Perceiving and Modes of Acting”. *In*: PICK, Herbert L., et SALTZMAN, Elliot (Eds.), *Modes of Perceiving and Processing Information*, Lawrence Erlbaum Associates, Hillsdale, N. J, 99-136.

TREVARTHEN, Colwyn (1984). “Biodynamic Structures, Cognitive Correlates of Motive Sets and the Development of Motives in Infants”. *In:* PRINZ, Wolfgang et SAUNDERS, A.F. *Cognition and Motor Processes*, Springer-Verlag, Berlin, 327-350.

TREVARTHEN, Colwyn and SPERRY, Roger Wolcott. (1973) “Perceptual Unity of Ambient Visual Field in Human Commissurotomy Patients”. *Brain*, 96 : 547-570.

TVERSKY, Amos and KAHNEMAN, Daniel (1981). “The framing of decisions and the rationality of choice”, *Science*, 211:453-458.

VAN INWAGEN, Peter (1978). “The Possibility of Resurrection”. *International Journal for the Philosophy of Religion*, 9:114-121.

VON NEURATH, Otto (1913). “Radical Physicalism and the Real World”. *Philosophical Papers*.

WILLIAMS, Bernard. (1970). “The self and the future”, *In:* WILLIAMS, Bernard (1973). *Problems of the Self*. Cambridge University Press, Cambridge, 1976, 274 p.

YABLO, Stephen (2002). “Coulda, Woulda, Shoulda.”. *In:* GENDLER, Tamar Szabò et HAWTHORNE, John (2002). *Conceivability and Possibility*. Clarendon Press, Oxford, p. 441-492.

## II) Supports médiagraphiques (Articles, documents, œuvres, disponibles en ligne, sur la toile) :

BANACH, David (1992). “Who do you think you are? Relations, Subjectivity, and the Identity of Persons.” *Proceedings of the American Catholic Philosophical Association*, 66. [en ligne].

Disponible sur : <http://www.anselm.edu/homepage/dbanach/who.htm>.

BERGSON, Henri (1959). *L'Evolution créatrice*. Presses Universitaires de France, Paris, 374 p. [en ligne]

Disponible sur : <http://classiques.uqac.ca/classiques/bergson_henri/evolution_creatrice/evolution_creatrice.pdf>

BOSTROM, Nick (2003). Are you living in a computer simulation? *Philosophical Quarterly*, 53 (N°211) : 243-255. [en ligne]

Disponible sur : <http://www.simulation-argument.com>

BOULENGER, Abbé Auguste. La doctrine catholique. Paris-Lyon, éd. Emmanuel Vitte, 1927, Tome 1 : Le Dogme, 197 p. [En ligne]

Disponible sur : <http://www.jesusmarie.com/auguste_boulenger_precis_doctrine_chretienne.pdf>

BRICMONT, Jean (1999). Science et religion : l'irréductible antagonisme. Revue de l'Université de Bruxelles, Complexes n°1. [en ligne]

Disponible sur : <http://dogma.free.fr/txt/JB-Science01.htm>

BROWN, James Robert (2006) "Thought Experiments" *In: Stanford Encyclopedia of Philosophy* [en ligne].

Disponible sur : <http://plato.stanford.edu/entries/thought-experiment/>

BROWN, Mark T. (1999). "Two Advantages of Punctualist Theories of Personal Identity", *Mountain Plains Philosophy Conference*. [Conférence] [en ligne]

Disponible sur : <http://www.manitowoc.uwc.edu/staff/awhite/mb98.htm>

CHALMERS, David John (2002). "Does Conceivability Entail Possibility?", *In:* GENDLER, Tamar Szabò and HAWTHORNE, John (2002). *Conceivability and Possibility*. Clarendon Press, Oxford, p. 145-200. [en ligne]

Disponible sur : <http://consc.net/papers/conceivability.html>

CHALMERS, David John (2005). « The Matrix as Metaphysics". *In:* GRAU, Christopher, Philosophers Explore the Matrix, Oxford University Press, Oxford, 352 p. [en ligne]

Disponible sur : <http://consc.net/papers/matrix.pdf>

CHAPUIS, Delphine. « Les sense data jouent-ils un rôle dans la connaissance ? ». *In: Les sense-data. Histoire et actualité*. Amiens, 13-14 mars 2003. [Conférence] [en ligne]

Disponible sur : <http://sensedata.free.fr/txt2003/DC2003.pdf>

CLARK, Andy and CHALMERS, David John (1998). "The Extended Mind". *Analysis*, 58 : 10-23. [en ligne]

Disponible sur : < http://consc.net/papers/extended.html>

CLARK, Thomas W. (1994). Death, Nothingness, and Subjectivity. *The Humanist*, 54(4) : 15-21. [en ligne]

Disponible sur : <http://www.naturalism.org/death.htm>

CLEEREMANS, Axel (2005). "L'Unité de la conscience". *In* : CAZENAVE, Michel (Ed.) *De la science à la philosophie : Y a-t-il une unité de la connaissance ?*, 147-171.

Disponible sur : <http://srsc.ulb.ac.be/axcWWW/papers/pdf/03-Unity.pdf>
Page consultée le 20 septembre 2006.

DELARUE, Michel (2000). « "De l'oeuf à la grenouille". Les Etapes du Développement d'un Amphibien : Le xenope (Xenopus laevis) » [en ligne]

Disponible sur : <http://www.snv.jussieu.fr/bmedia/xenope1/index.htm>
Mise à jour : juin 2003. Page consultée le 20 septembre 2006.

DENNETT, Daniel Clement (1990). "Quining Qualia". *In:* LYCAN, William G. *Mind and Cognition.* Blackwells, Oxford, p. 519-548.

Disponible sur : <http://ase.tufts.edu/cogstud/papers/quinqual.htm>

DOKIC, Jérôme (1998) La perception interne et la critique du langage privé. *Revue de théologie et de philosophie*, 130 : 1-19.

Disponible sur :
<http://jeannicod.ccsd.cnrs.fr/docs/00/05/33/93/PDF/ijn_00000275_00.pdf>

FOUERE, René (1969). *Krishnamurti ou la révolution du réel*. Le Courrier du Livre, Paris, 1993, 446 p. [en ligne]

Disponible sur : <http://www.lequasar.net/fouererevolution.pdf>

HEYLIGHEN, Francis, JOSLYN, Cliff et TURCHIN, Valentin, *Principia Cybernetica Web* [en ligne]

Disponible sur : <http://pespmc1.vub.ac.be/>
Mis à jour le 28 février 2006, créé le 25 octobre 1993.

HUEMER, Michael, A Proof of Free Will [en ligne]

Disponible sur : <http://home.sprynet.com/~owl1/fwill.htm>
Page consultée le 20 septembre 2006.

HURLEY Susan (2002). "Action, the unity of consciousness, and vehicle externalism". *In:* Cleeremans, Axel. *The Unity of Consciousness: Binding, Integration, Dissociation*. Oxford University Press, Oxford. [en ligne]

Disponible sur Internet :
<http://www.warwick.ac.uk/staff/S.L.Hurley/papers/auc.pdf>

GENDLER, Tamar Szabò (1998). "Galileo and the Indispensability of Scientific Thought Experiment". *The British Journal for the Philosophy of Science*, 49/3 (Sept): 397-424 [en ligne]

Disponible sur : <http://bjps.oxfordjournals.org/cgi/reprint/49/3/397.pdf>

GENDLER, Tamar Szabò (1998). "Exceptional Persons: On the Limits of Imaginary Cases". *Journal of Consciousness Studies*, 5:592-610. [en ligne]

Disponible sur :
<http://www.people.cornell.edu/pages/tsg3/exceptionalpersons.pdf>

GENDLER, Tamar Szabò (2002). "Personal Identity and Thought-Experiments". *Philosophical Quarterly*, 52/206:34-54. [en ligne]

Disponible sur : <http://www.people.cornell.edu/pages/tsg3/personalidentity.pdf>

JACKSON, Frank (1982). "Epiphenomenal Qualia". *Philosophical Quarterly*, 32:27-36 [en ligne]

Disponible sur : <http://members.aol.com/NeoNoetics/Mary.html>

KAPLAN, Francis (2004) L'irréalité de la mort. *In:* LENOIR, Frédéric et DE TONNAC, Jean-Philippe. La mort et l'immortalité : Encyclopédie des savoirs et croyances. Bayard, Paris, p. 1424-1449. [en ligne]

Disponible sur : <http://www.dtext.com/francis-kaplan/irrealite-mort.html>

LLOYD, Peter B. (1993). "Is the Mind Physical? Dissecting Conscious Brain Tissue", *Philosophy Now*, No. 6. [en ligne]

Disponible sur : <http://www.newdualism.org/papers/P.LLoyd/phinow2.htm>

MAGNAN, Christian, « L'infini : ça n'existe pas ! » [en ligne]

Disponible sur : <http://www.lacosmo.com/infini.html>

MAGNAN, Christian, « L'infini des cosmologistes : réalité ou imposture ? » [en ligne],

Disponible sur : <http://www.lacosmo.com/infini-encore.html>

MARCHAL, Bruno (1998). Calculabilité, Physique et Cognition. Lille. Thèse de doctorat

en informatique.

Disponible sur : <http://iridia.ulb.ac.be/~marchal>

MEIXNER, Uwe (2006). On Some Realisms Most Realists Don't Like. Colloque International (Anti-)Réalisme : Logique et Métaphysique, 28 juin-1er juillet 2006, Université Nancy2, 30 juin 2006. [Conférence.]

NORTON, John (2004). "Why Thought Experiments Do Not Transcend Empiricism", *In:* HITCHCOCK, Christopher (2004). *Contemporary Debates in the Philosophy of Science*. Blackwell, Oxford, p. 44-66. [en ligne]

Disponible sur : <http://philsci-archive.pitt.edu/archive/00000960/>

NORTON, John (2004). "On Thought Experiments: Is There More to the Argument?". *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science*, 71:1139-1151. [en ligne]

Disponible sur :
<http://www.pitt.edu/~jdnorton/papers/PSA_2002_te.pdf>

O'REGAN, John Kevin et NOË, Alva. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24 : 939-1031. [en ligne]

Disponible sur Internet :
<http://nivea.psycho.univ-paris5.fr/OREGAN-NOE-BBS/ORegan;Noe.BBS.pdf>

PLANTINGA, Alvin (2003). An evolutionary argument against naturalism. *In* : Faith in Theory and Practice: Essays on Justifying Religious Belief. Open Court Publishing Company, Chicago, Illinois. [en ligne]

Disponible sur : <http://hisdefense.org/articles/ap001.html>

RAYMONT, Paul. "The Unity of Consciousness". *In: Stanford Encyclopedia of Philosophy*. [en ligne]

Disponible sur : <http://plato.stanford.edu/entries/consciousness-unity/>

RESCHER, Nicholas. (2002) Process Philosophy. *In: Stanford Encyclopedia of Philosophy*.

Disponible sur : <http://plato.stanford.edu/entries/process-philosophy/>]

ROSAT, Jean-Jacques (2001). « D'un bon usage des sense-data ». *In : Les sense-data. Histoire et actualité*. Amiens, 13-14 décembre 2001. [Conférence] [en ligne]

Disponible sur : <http://sensedata.free.fr/JJRosat/JJRosat.htm>

STEWART, Wayne (2006). *Metaphysics by default* [en ligne]

Disponible sur : <http://mbdefault.org>

STONEHAM, Tom (2003). Temporal Externalism. *Philosophical Papers*, 32 : 97-107.

Disponible sur : <http://www-users.york.ac.uk/ twcs1/Papers/Temporal.pdf>

VEZEANU, Ion (2005). « "Ce doit être moi puisque je suis ici !". Agnosie et reconnaissance ». *Studia, collection Philosophia, Universitatae Babes-Bolyai*, L(2) : 153-178. [en ligne]

Disponible sur :
<http://web.upmf-grenoble.fr/SH/PersoPhilo/IonVezeanu/AgnoReconWeb.pdf>

SINGER, Wolf (2001). "Consciousness and the binding problem". *In:* MARIJUÁN, Pedro C., Cajal and Consciousness: Scientific Approaches to Consciousness on the Centennial of Ramòn y Cajal's Textura. *Annals of the New York Academy of Science*, 929:123-146. [en ligne]

Disponible sur :
<http://www.mpih-frankfurt.mpg.de/global/Np/Pubs/zaragoza.pdf>

ZWIRN, Hervé (2002). « Mécanique quantique et connaissance du réel ». *In :* D'ESPAGNAT, Bernard (Ed.) (2002). *Implications philosophiques de la science contemporaine*, Presses Universitaires de France, Vol. 2., 76 p.

Disponible sur : <http://www.asmp.fr/travaux/gpw/philosc/rapport2/5-Zwirn.pdf>

# _INDEX_


## A

Accident, 206, 412
Agnosticisme, 146, 381
Âme, 8, 21, 28, 30, 50, 51, 67, 77, 101, 111, 119, 125, 126, 127, 157, 161, 178, 219, 220, 223, 224, 227, 262, 281, 287, 299, 440, 459, 469, 492, 496, 512, 532, 533, 541, 558
Antiréductionnisme, 23, 93, 109, 110, 115, 469
Athéisme, 25, 438

## B

Bergson, 454, 457, 499, 500, 502, 508, 525, 560
Blackmore, 523, 539, 552
Brown, 37, 40, 41, 44, 91, 553, 557, 561

## C

Carnap, 31, 251, 268, 552
Cas H.M., 207, 208, 209, 210, 212, 213, 216, 217
Cas Wearing, 213, 216, 217, 556
Cerveau, 65, 68, 74, 84, 87, 89, 90, 94, 102, 103, 118, 130, 194, 201, 203, 205, 207, 215, 252, 265, 266, 283, 285, 291, 293, 295, 297, 299, 300, 302, 303, 306, 307, 308, 309, 311, 321, 324, 326, 335, 338, 343, 355, 357, 385, 408, 412, 440, 444, 453, 474, 475, 498, 507, 508, 509, 512, 519, 535, 539, 540, 553
  hémisphères, 84, 86, 87, 88, 89, 194, 195, 196, 201, 203, 204, 296, 297, 298, 306, 307, 316, 317, 332, 337, 412, 452, 480, 482
  planétaire, 74, 308, 309, 311, 512
Chalmers, 8, 280, 290, 291, 292, 293, 360, 362, 366, 367, 368, 369, 372, 561
Changement, 34, 61, 69, 81, 86, 118, 189, 190, 192, 196, 206, 212, 303, 312, 320, 321, 327, 340, 458, 462, 473, 504, 505, 515
Chauvier, 45, 170, 552
Chose, 16, 18, 19, 34, 37, 44, 45, 50, 55, 61, 66, 67, 77, 80, 81, 82, 86, 89, 91, 93, 98, 99, 100, 101, 103, 119, 121, 122, 125, 130, 132, 135, 138, 140, 141, 143, 153, 157, 176, 181, 184, 198, 209, 219, 224, 225, 227, 237, 240, 242, 243, 245, 249, 251, 254, 259, 262, 266, 270, 278, 279, 280, 283, 285, 288, 302, 303, 306, 311, 313, 315, 319, 322, 323, 331, 343, 350, 351, 352, 353, 363, 369, 378, 379, 384, 387, 389, 391, 399, 400, 404, 407, 418, 419, 424, 425, 437, 440, 446, 449, 450, 457, 461, 493, 494, 495, 498, 499, 500, 502, 503, 504, 505, 508, 514, 519, 520, 525, 526, 532
Circularité, 165, 166, 170, 404
Clark, Andy, 290, 561
Clark, Thomas W., 528
Cognition, 141, 285, 286, 291, 475
Connexité, 7, 23, 61, 62, 85, 168, 169, 183, 186, 188, 201, 207, 208, 209, 210, 212, 213, 215, 216, 224, 535, 536
Conscience, 8, 17, 23, 24, 25, 37, 43, 45, 51, 52, 53, 54, 55, 56, 57, 58, 60, 61, 65, 66, 68, 69, 72, 74, 83, 84, 85, 86, 87, 88, 89, 92, 93, 114, 118, 119, 129, 130, 131, 132, 154, 155, 167, 172, 196, 199, 205, 206, 208, 213, 215, 216, 217, 219, 220, 221, 223, 224, 226, 230, 241, 243, 246, 249, 250, 265, 267, 269, 274, 277, 280, 285, 288, 293, 295, 297, 302, 303, 306, 307, 308, 313, 317, 319, 322, 323, 326, 330, 331, 332, 333, 334, 335, 337, 338, 339, 340, 341, 342, 343, 389, 441, 442, 447, 455, 471, 472, 473, 481, 482, 484, 485, 494, 498, 499, 503, 504, 505, 506, 512, 514, 517, 518, 519, 525, 526, 529, 530, 536, 537, 540, 541, 553, 562
  de soi, 72, 83, 130, 172, 205, 471, 472, 473, 481, 484, 485, 536, 541
    en première personne (conscience égologique), 130, 471, 472, 473, 481, 484, 485, 536, 541
Continuité, 7, 16, 23, 61, 71, 73, 85, 92, 98, 101, 106, 109, 111, 113, 115, 168, 169, 173, 178, 183, 186, 201, 206, 207, 208, 209, 218, 220, 224, 226, 311, 389, 391, 404, 407, 409, 419, 461, 473, 476, 477, 478, 479, 480, 481, 482, 492, 514, 521, 525, 526, 529, 530, 536, 538, 540, 541

## D

Damasio, 55, 215, 217, 552
Déisme, 282
Dennett, 19, 41, 42, 52, 65, 244, 246, 254, 265, 297, 299, 301, 335, 355, 356, 539, 553, 558, 562
Descartes, 50, 65, 227, 240, 267, 281, 295, 326, 454, 531, 553
Déterminisme, 41, 125, 132, 133, 134, 137, 140, 145, 148, 149, 150, 151, 152, 200, 204, 221, 329, 376, 434
Dieu, 21, 25, 26, 27, 28, 30, 32, 47, 110, 151, 212, 222, 226, 281, 282, 309, 368, 374, 437, 438, 447, 448, 456, 467, 496, 497, 542, 551
Dualisme, 66, 119, 122, 151, 264, 281, 371, 512

## E

Eccles, 66, 448, 531, 553, 555
Ego, 33, 121, 122, 126, 192, 471, 523, 524, 531
Eliminativisme, 65
Emotion, 25, 114, 131, 213, 304, 495, 504, 541
Empirie, 19, 40, 68, 99, 107, 110, 111, 113, 156, 265, 381, 433, 469, 484, 496
Empirisme, 39, 40, 267, 454
Epistémologie, 7, 25, 27, 29, 38, 134, 399, 436, 475
Esprit, 3, 8, 15, 19, 24, 32, 35, 54, 59, 65, 68, 72, 74, 80, 81, 88, 102, 133, 134, 143, 150, 151, 160, 161,

162, 167, 182, 185, 189, 197, 206, 214, 218, 222, 249, 250, 278, 281, 282, 283, 285, 286, 290, 293, 298, 299, 305, 306, 308, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 330, 332, 372, 376, 378, 383, 387, 394, 398, 428, 433, 451, 456, 461, 474, 485, 497, 500, 503, 512, 515, 517, 555, 556, 558
Essence, 28, 45, 48, 68, 72, 86, 100, 101, 310, 311, 401, 541
Etat, 9, 18, 34, 41, 45, 85, 88, 113, 130, 133, 137, 138, 189, 206, 215, 222, 226, 241, 269, 283, 284, 303, 309, 326, 331, 381, 408, 409, 433, 439, 447, 455, 456, 461, 463, 498, 499, 502, 503, 504, 505, 514, 526, 528, 539
Eternité, 9, 18, 441, 489, 509, 519, 522, 544
Être, 3, 7, 9, 13, 15, 16, 17, 20, 21, 23, 25, 27, 28, 30, 31, 32, 35, 36, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 52, 54, 55, 57, 58, 59, 60, 61, 63, 66, 67, 68, 70, 71, 73, 74, 77, 79, 80, 81, 82, 84, 85, 87, 88, 89, 90, 91, 93, 95, 97, 98, 99, 101, 102, 103, 105, 106, 107, 108, 109, 110, 112, 113, 114, 115, 116, 118, 123, 124, 125, 126, 127, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 145, 146, 147, 148, 149, 151, 152, 154, 155, 156, 157, 158, 160, 161, 162, 163, 166, 167, 170, 172, 173, 174, 175, 177, 179, 180, 181, 183, 184, 185, 188, 190, 191, 192, 193, 194, 196, 198, 202, 203, 205, 210, 212, 213, 216, 218, 219, 220, 221, 222, 223, 224, 225, 227, 228, 229, 230, 232, 235, 236, 238, 239, 240, 241, 242, 243, 246, 248, 250, 251, 252, 253, 254, 258, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 272, 273, 274, 277, 279, 280, 281, 282, 283, 285, 286, 287, 288, 289, 290, 292, 294, 295, 296, 297, 299, 302, 303, 305, 306, 308, 309, 310, 311, 312, 314, 315, 316, 317, 319, 320, 322, 325, 326, 327, 328, 330, 331, 332, 333, 335, 338, 340, 342, 351, 352, 353, 354, 355, 356, 357, 358, 359, 361, 363, 364, 365, 366, 368, 370, 371, 372, 373, 376, 377, 378, 379, 380, 381, 383, 384, 386, 387, 390, 391, 394, 398, 399, 401, 403, 404, 405, 408, 410, 411, 417, 418, 420, 421, 422, 425, 431, 433, 435, 436, 437, 438, 439, 440, 443, 444, 445, 447, 449, 450, 452, 453, 456, 457, 458, 459, 460, 461, 462, 463, 467, 469, 471, 472, 473, 474, 475, 476, 477, 480, 482, 483, 484, 485, 486, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 507, 508, 509, 510, 513, 515, 516, 518, 520, 521, 523, 525, 526, 527, 529, 531, 533, 537, 538, 542, 543, 549, 558, 565
Evénement, 85, 95, 100, 133, 165, 166, 170, 186, 193, 213, 304, 327, 463, 489, 515, 516, 518, 521
Expérience, 23, 24, 27, 31, 37, 38, 39, 40, 41, 42, 43, 51, 52, 53, 54, 55, 56, 58, 84, 86, 91, 94, 95, 96, 104, 109, 112, 119, 124, 127, 128, 130, 133, 143, 144, 145, 153, 154, 155, 156, 176, 179, 180, 181, 188, 189, 190, 194, 196, 197, 203, 205, 206, 211, 213, 216, 235, 240, 246, 249, 250, 251, 253, 254, 264, 267, 274, 275, 276, 277, 284, 287, 289, 292, 293, 295, 299, 312, 317, 319, 322, 349, 351, 352, 353, 356, 378, 379, 382, 385, 387, 390, 397, 402, 403, 404, 405, 406, 409, 411, 412, 421, 428, 431, 433, 439, 447, 451, 456, 461, 463, 481, 485, 491, 498, 505, 517, 518, 523, 525, 526, 527, 528, 529, 530, 537
Externalisme, 7, 8, 24, 73, 74, 102, 155, 227, 228, 283, 284, 285, 287, 288, 290, 295, 299, 300, 310, 313, 315, 328, 329, 330, 338, 481, 482, 512, 519, 540

## F

Ferret, 49, 72, 80, 81, 93, 103, 224, 553

## G

Gendler, 8, 39, 350, 360, 369, 378, 381, 382, 383, 385, 386, 387, 388, 389, 390, 391, 392, 393, 397, 402, 406, 407, 410, 411, 412, 413, 414, 415, 416, 418, 419, 420, 421, 423, 425, 426, 427, 428, 429, 430, 431, 432, 433, 560, 561, 563

## H

Hurley, 331, 332, 333, 337, 338, 339, 340, 562, 563
Husserl, 58, 267, 553

## I

Identité, 1, 7, 8, 9, 15, 19, 20, 21, 22, 23, 24, 35, 36, 42, 44, 46, 47, 48, 49, 50, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 79, 80, 81, 82, 83, 84, 85, 90, 91, 92, 93, 94, 96, 98, 99, 100, 101, 102, 104, 109, 110, 112, 114, 115, 117, 118, 119, 122, 126, 129, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 170, 172, 173, 174, 175, 176, 177, 180, 181, 183, 184, 186, 189, 195, 197, 200, 201, 202, 203, 207, 216, 218, 220, 222, 223, 224, 225, 226, 227, 228, 230, 231, 232, 236, 238, 307, 309, 310, 311, 313, 315, 323, 329, 332, 342, 343, 349, 350, 351, 360, 366, 378, 381, 383, 385, 389, 390, 391, 392, 393, 395, 396, 397, 404, 407, 409, 410, 413, 414, 415, 419, 420, 421, 422, 423, 424, 425, 426, 429, 430, 431, 432, 433, 435, 438, 439, 440, 441, 442, 447, 450, 451, 452, 459, 462, 463, 464, 467, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 482, 483, 484, 491, 492, 493, 509, 512, 513, 519, 522, 523, 531, 532, 534, 535, 536, 538, 541, 542, 543, 544, 545, 553
numérique, 7, 23, 47, 61, 62, 63, 68, 69, 70, 71, 74, 79, 80, 85, 114, 165, 167, 170, 172, 173, 180, 181, 183, 195, 201, 203, 218, 227, 231, 232, 287, 307, 462, 477, 480, 522, 538
personnelle, 1, 7, 8, 9, 15, 19, 20, 21, 22, 23, 24, 35, 36, 42, 46, 47, 48, 49, 50, 63, 64, 65, 66, 67, 68, 69, 70, 73, 74, 79, 80, 81, 83, 84, 91, 92, 93, 94, 96, 98, 99, 100, 109, 110, 112, 115, 117, 118, 119, 122, 126, 129, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 170, 172, 173, 174, 175, 177, 180, 181, 183, 184, 186, 189, 200, 202, 203, 207, 216, 218, 220, 223, 224, 225, 226, 227, 228, 236, 238, 310, 312, 313, 315, 323, 329, 342, 343, 349, 350, 360, 378, 381, 383, 385, 389, 390, 391, 392, 393, 397, 404, 407, 409, 410, 424, 431, 432, 433, 435, 438, 439, 440, 441, 442, 447, 450, 451, 452, 459, 464, 469, 470, 471, 474, 475, 476, 477, 478, 479, 482, 483, 484, 491,

492, 493, 512, 513, 519, 522, 523, 531, 532, 534, 535, 536, 538, 541, 542, 544, 545
qualitative, 7, 23, 54, 61, 62, 73, 126, 168, 196, 241
spécifique, 28, 43, 263, 294, 522
Identité diachronique, 23, 56, 57, 58, 61, 85, 213, 391, 392, 393, 482, 484, 522
Identité synchronique, 23, 56, 522
Immortalité, 18, 50, 299, 440, 492, 494, 496, 524, 563
Indéterminisme, 88, 149, 178, 199, 204, 329, 330, 447
Individu, 23, 36, 44, 46, 47, 48, 55, 62, 72, 73, 83, 91, 99, 103, 106, 107, 117, 120, 123, 125, 126, 127, 136, 152, 156, 162, 168, 172, 175, 178, 185, 187, 188, 191, 192, 194, 195, 197, 201, 204, 205, 210, 212, 218, 220, 222, 225, 227, 228, 251, 260, 262, 285, 294, 298, 307, 309, 310, 311, 312, 319, 320, 321, 330, 409, 412, 434, 439, 452, 467, 469, 472, 473, 474, 476, 477, 480, 481, 521, 528, 537, 541
Individualité, 184, 185, 312, 537
Internalisme, 8, 24, 155, 228, 283, 293, 294, 300, 324, 481
Intersubjectivité, 258, 447
Intuition, 21, 131, 132, 139, 145, 146, 153, 156, 157, 159, 198, 244, 250, 281, 283, 286, 338, 352, 386, 387, 446, 491, 524, 526, 529
Inwagen, 72, 149, 556, 560

## J

Je, 459

## K

Kant, 30, 49, 65, 103, 138, 139, 267, 275, 453, 454, 458, 531, 538, 554
Kaplan, 59, 60, 214, 516, 518, 519, 555, 563

## L

Leibniz, 69, 110, 162, 163, 216, 218, 220, 221, 222, 223, 225, 226, 252, 523, 554
Lewis, 13, 175, 192, 204, 224, 551, 558
Liberté, 17, 132, 135, 138, 141, 149, 150, 151, 152, 153, 159, 221, 327, 329
Logique, 9, 20, 30, 32, 35, 39, 40, 61, 82, 85, 86, 98, 103, 143, 145, 173, 174, 175, 177, 183, 184, 191, 199, 202, 203, 208, 210, 251, 252, 253, 268, 279, 282, 294, 300, 304, 313, 324, 351, 354, 370, 371, 403, 404, 454, 494, 496, 498, 525, 544, 552
Lucrèce, 37, 42, 70, 71, 554

## M

Mach, 20, 39, 554
Marchal, 178, 191, 281, 563
Matérialisme, 298, 353, 354, 356, 380, 496, 498
Matière, 17, 18, 31, 50, 65, 68, 69, 70, 93, 109, 110, 113, 115, 150, 151, 154, 157, 159, 161, 171, 201, 222, 226, 237, 238, 239, 281, 287, 309, 326, 405, 406, 408, 441, 470, 473, 482, 483, 504, 521
Mémoire, 16, 47, 59, 60, 70, 85, 91, 93, 94, 101, 111, 114, 159, 160, 162, 165, 166, 170, 185, 195, 201, 205, 207, 208, 214, 215, 216, 217, 219, 221, 276, 292, 293, 298, 332, 354, 383, 387, 451, 457, 503, 512, 517, 518, 519, 520, 521, 526, 528, 531, 540, 541
Métaphysique, 20, 21, 22, 28, 29, 31, 32, 35, 36, 37, 106, 110, 111, 112, 113, 115, 118, 133, 135, 136, 137, 146, 150, 158, 179, 238, 252, 262, 267, 275, 277, 280, 287, 288, 315, 324, 325, 359, 360, 361, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 380, 381, 409, 420, 421, 422, 426, 432, 433, 434, 435, 436, 459, 470, 486, 491, 496, 498, 506, 542, 545
Milner, 207, 559
Moi, 49, 93, 94, 105, 111, 162, 207, 289, 469, 470, 499, 537
Monde actuel, 138, 351, 475, 484
Monde possible, 138
Mort, 16, 59, 60, 66, 70, 73, 82, 90, 91, 155, 157, 186, 188, 193, 199, 210, 212, 214, 298, 313, 342, 433, 440, 467, 489, 492, 493, 494, 495, 498, 501, 502, 504, 505, 507, 508, 509, 510, 512, 513, 514, 515, 516, 520, 521, 522, 523, 524, 525, 526, 529, 530, 534, 537, 538, 563

## N

Nagel, 54, 120, 240, 252, 316, 317, 318, 319, 320, 321, 322, 323, 331, 332, 459, 523, 558, 559
Néant, 7, 9, 18, 66, 155, 321, 326, 441, 454, 457, 460, 486, 493, 494, 495, 496, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 513, 514, 515, 525, 526, 530, 537, 544, 545
Neurath, 251, 560
Norton, 38, 40, 553, 564
Nozick, 177, 180, 181, 224, 389, 554, 555

## O

Objectivité, 32, 50, 65, 68, 95, 99, 103, 120, 129, 131, 158, 190, 227, 249, 311, 380, 443, 446, 447, 450, 454, 498, 503, 506, 514, 529
Objet, 28, 32, 40, 45, 49, 54, 56, 58, 59, 104, 118, 152, 153, 158, 160, 176, 180, 240, 241, 242, 249, 250, 253, 254, 257, 258, 259, 260, 265, 268, 269, 271, 273, 283, 287, 288, 289, 303, 310, 333, 352, 381, 405, 406, 434, 457, 496, 499, 500, 503, 508, 518
Olson, 72, 555
Ontologie, 7, 22, 23, 24, 25, 27, 28, 29, 35, 128, 174, 193, 198, 202, 227, 280, 315, 325, 327, 330, 343, 436, 475, 481, 482, 535

## P

Parfit, 8, 55, 62, 66, 72, 84, 85, 86, 89, 90, 91, 97, 165, 168, 170, 171, 183, 185, 188, 191, 194, 201, 203, 207, 208, 209, 210, 216, 224, 238, 350, 407, 411, 412, 413, 415, 419, 420, 421, 422, 423, 424, 426, 427, 428, 429, 430, 431, 433, 453, 523, 525, 555, 559
Perry, 175, 192, 204, 224, 316, 559

Personnalité, 16, 46, 68, 103, 107, 136, 185, 194, 195, 205, 225, 227, 336, 342, 383, 396, 451, 453, 472, 473, 520, 526, 528, 533, 541
Personne, 7, 17, 18, 19, 22, 23, 24, 26, 33, 35, 36, 44, 45, 46, 47, 48, 49, 50, 55, 63, 67, 68, 70, 71, 72, 79, 80, 81, 82, 83, 85, 90, 93, 94, 95, 96, 97, 101, 103, 106, 107, 109, 110, 111, 112, 113, 114, 118, 122, 123, 124, 129, 130, 131, 135, 137, 139, 142, 144, 145, 147, 148, 155, 158, 161, 162, 165, 166, 167, 169, 170, 171, 172, 173, 174, 175, 177, 181, 182, 183, 184, 185, 188, 190, 191, 192, 193, 194, 195, 197, 198, 201, 202, 203, 204, 205, 207, 210, 212, 220, 225, 226, 227, 228, 229, 230, 231, 232, 238, 241, 243, 244, 247, 253, 254, 260, 261, 262, 263, 274, 275, 277, 281, 282, 285, 286, 290, 291, 292, 294, 295, 296, 297, 298, 303, 311, 312, 315, 318, 319, 320, 321, 322, 323, 327, 330, 335, 336, 342, 343, 351, 352, 355, 372, 381, 382, 383, 385, 386, 387, 388, 389, 391, 392, 393, 396, 401, 406, 407, 408, 410, 417, 432, 433, 440, 441, 442, 449, 451, 452, 453, 467, 470, 471, 472, 473, 474, 475, 476, 477, 479, 480, 482, 483, 484, 485, 486, 492, 494, 495, 496, 498, 505, 506, 507, 508, 513, 519, 520, 523, 524, 526, 527, 528, 530, 531, 534, 535, 539, 540, 541, 542, 543, 552
Physicalisme, 280, 293
Plantinga, 144, 146, 380, 381, 555, 564
Popper, 66, 448, 555
Pouivet, 1, 3, 25, 72, 98, 163, 164, 272, 278, 555, 559
Processus, 22, 27, 29, 30, 69, 71, 73, 80, 85, 88, 93, 97, 106, 107, 108, 109, 130, 132, 138, 154, 156, 160, 180, 187, 188, 190, 191, 195, 196, 230, 234, 249, 250, 254, 286, 289, 291, 300, 313, 315, 316, 325, 327, 328, 329, 330, 337, 343, 352, 354, 393, 403, 405, 423, 424, 447, 455, 474, 475, 476, 477, 480, 481, 482, 492, 504, 521, 527, 529, 538, 539
Propriété, 94, 113, 131, 178, 232, 275, 371, 372, 406, 458, 502, 504

## Q

Qualia, 42, 43, 54, 244, 245, 246, 252, 262, 264, 269, 273, 277, 285, 287, 288, 293, 355, 358

## R

Raison, 29, 63, 67, 69, 83, 86, 102, 129, 136, 139, 141, 142, 143, 144, 145, 154, 156, 159, 162, 163, 176, 185, 193, 201, 212, 221, 246, 258, 299, 301, 359, 371, 421, 437, 446, 447, 448, 492, 493, 496, 506, 513, 520, 525, 554
Réductionnisme, 23, 93, 295, 469
Réincarnation, 222, 396, 512
Ressemblance, 7, 16, 41, 61, 73, 83, 121, 159, 160, 178, 184, 189, 225, 276, 319, 353, 367, 382, 418, 426
Résurrection, 230
Rosnay (De), Joël, 74, 308, 309, 540, 553
Russell, 214, 247, 249, 251, 555, 556, 559
Ryle, 227, 282, 283, 295, 358, 556

## S

Schlick, 250, 251, 559
Sciences cognitives, 1, 3, 23, 27, 35, 326
Scoville, 207, 559
Siamois (jumeaux), 8, 48, 57, 126, 128, 170, 194, 315, 341, 342, 472, 474, 477
Soi, 8, 16, 29, 30, 68, 71, 92, 130, 135, 140, 141, 175, 195, 215, 216, 217, 221, 226, 228, 237, 238, 239, 241, 254, 267, 274, 277, 305, 311, 349, 383, 385, 395, 415, 423, 429, 438, 442, 448, 449, 453, 454, 455, 458, 459, 460, 462, 463, 464, 467, 471, 472, 473, 483, 484, 485, 486, 491, 493, 497, 498, 506, 533, 535, 545, 552
  témoin, 81, 166, 170, 453, 454, 455, 458, 459, 460, 463, 464, 483, 484, 485, 486, 498, 533, 545
Sousa (De), Ronald, 224, 225, 559
Souvenirs, 59, 60, 67, 68, 83, 96, 103, 106, 107, 108, 114, 117, 207, 212, 213, 216, 220, 227, 229, 387, 455, 461, 472, 473, 485, 517, 520, 527, 528, 541
Spinoza, 132, 556
Split-Brain (patients), 24, 48, 295, 316
Subjectivité, 32, 50, 86, 183, 227, 241, 243, 258, 264, 377, 434, 454, 492, 515, 526, 528, 530, 538
Substance, 45, 51, 68, 69, 70, 111, 128, 150, 157, 220, 228, 284, 316, 324, 328, 332, 363, 367, 481, 482
Sujet, 9, 17, 21, 24, 29, 45, 50, 52, 54, 65, 79, 93, 94, 99, 101, 104, 114, 120, 131, 132, 136, 137, 147, 149, 150, 155, 157, 158, 159, 165, 170, 183, 201, 215, 223, 229, 238, 240, 243, 247, 250, 253, 257, 258, 259, 260, 261, 263, 269, 271, 272, 278, 281, 286, 287, 289, 292, 294, 310, 319, 322, 328, 331, 333, 335, 343, 349, 350, 352, 353, 356, 358, 363, 364, 367, 371, 372, 373, 374, 376, 377, 379, 383, 387, 389, 390, 391, 397, 407, 408, 424, 427, 432, 433, 442, 445, 447, 449, 450, 454, 455, 459, 482, 500, 519, 520, 522, 525, 529, 530, 544, 549
Survie, 9, 24, 85, 90, 91, 141, 146, 155, 156, 182, 183, 186, 188, 190, 191, 195, 209, 210, 289, 298, 440, 444, 487, 492, 495, 519, 521, 522, 524, 525, 539, 544, 545

## T

Teilhard de Chardin, 74, 556
Télépathie, 87, 88, 123, 183
Téléportation, 182, 184, 205, 439, 473
Théisme, 146
Théologie, 22, 25, 27, 28, 45, 153, 163, 164, 467, 551, 552, 559, 562

## W

Wiggins, 72, 89, 90, 91, 98, 171, 181, 201, 556
Williams, 385, 386, 387, 389, 390, 391, 392, 393, 404, 407, 433, 560
Wittgenstein, 20, 21, 32, 33, 34, 163, 227, 239, 240, 242, 243, 246, 266, 271, 272, 273, 282, 283, 332, 351, 451, 452, 453, 454, 459, 489, 549, 555, 556, 557