Le secret de l'amibe

Bruno Marchal

19 mai 2000

À la mémoire de mes parents.

Table des matières

Chapitre 1

Introduction

L'esprit se retournera comme un gant.
Propos bouddhiste

Monsieur Edgar Morin, président du jury du *Prix de la Recherche Universitaire Le Monde 1998*, a recommandé aux lauréats dont je fais partie, et qui ont l'heureuse chance de voir leur thèse de doctorat publiée chez Grasset, de ne pas hésiter à présenter la thèse en tenant compte de l'aspect humain et personnel, ce qui m'a finalement décidé à raconter son histoire.

Je dois reconnaître le caractère un peu spécial du résultat obtenu, mais aussi et peut-être surtout, de la nature du chemin parcouru. J'ai exposé la première fois la question de base du travail en 1963, dans une école de la ville de Bruxelles. J'avais huit ans : je ne sais pas si j'étais particulièrement précoce ou simplement surangoissé, en effet la motivation du travail et de la recherche dès l'enfance, a toujours été directement liée à une inquiétude face à la mort.

Il existe d'autres raisons plus fondamentales pour décrire ce cheminement dont l'essentiel se situe dans la petite enfance.

1. Le travail est essentiellement interdisciplinaire. Il se situe à l'intersection de nombreuses disciplines—théologie, psychologie, biologie, chimie, physique, mathématiques, informatique—et, du coup, il est difficile de savoir par où commencer. À ce titre, le chemin des questions d'enfants est particulièrement bien adapté.
2. Je me suis posé des questions, mais rapidement je me suis demandé d'ou viennent ces questions et une bonne part de la thèse repose sur un processus d'auto-observation. On verra que la thèse est par nature auto-explicative, elle explique sa propre génèse. C'est un aspect qui est plus clair si on suit, même brièvement, le chemin des questions d'enfants. J'en profite aussi pour proposer une version plus travaillée

de l'argument principal. Il ne s'agit pas de vulgarisation et le lecteur peut sauter les passages qu'il juge trop techniques.

3. C'est l'occasion de rendre justice et hommage à de grands auteurs/livres qui ont jalonnés la quête parmi lesquelles G. Ames & R. Wyler, James Watson, Linus Pauling, Michel-Yves Bernard, Lewis Carroll, E. Nagel & J. R. Newman, Jean Ladrière, S. C. Kleene, Bernard d'Espagnat.
4. ... et de raconter une histoire *belge*, dans le meilleur des cas, ou *universelle* dans le pire des cas, qui n'est pas vraiment drôle. Elle explique cependant pourquoi je défendrai la thèse en 1998 en France. Je raconte ces évènements sans haine et sans esprit de revanche.

Je vous expose dès à présent le résultat principal en quelques mots. D'abord le travail présente une preuve, c'est-à-dire une argumentation déductive ou encore hypothético-déductive comme on dit plus lourdement. Cela signifie qu'il y a une hypothèse ainsi qu'une "thèse", dans le sens plus technique de ce qui est démontré à partir de l'hypothèse.

Par preuve on entend que si le lecteur n'est pas convaincu du résultat après avoir étudié le travail, il doit mettre en évidence soit une proposition non justifiée soit une erreur. Notons que dans un travail déductif les conclusions ne portent jamais *d'office* sur la réalité. Les preuves scientifiques opèrent dans le cadre d'une théorie qui est *postulée*. La science est ainsi toujours modeste au sujet de son applicabilité ou de son adéquation au réel.

L'hypothèse est celle du mécanisme : l'idée que nous pourrions être des machines numériques, dans un sens qui sera bien sûr précisé davantage. En gros nous serions des machines en ce sens qu'il n'existe pas de parties de notre corps qui soit privilégiée par rapport à une éventuelle substitution fonctionnelle : c'est-à-dire que l'on peut survivre à une substitution du cœur par un cœur artificiel, ou d'un rein par un rein artificiel, ou du cerveau par un cerveau artificiel, etc., pour autant que la substitution se fasse à un niveau suffisament fin. Il n'y aura par ailleurs aucune restriction sur la finesse du niveau imposée. Il importe de bien comprendre que je ne vais pas défendre l'hypothèse du mécanisme. Je vais seulement poser cette hypothèse au départ. Elle constitue le cadre prédéfini du travail[1].

[1]Notons que l'idée de prendre le mécanisme ou le computationnalisme comme hypothèse semble être, assez curieusement, originale. Depuis Descartes (et même avant, notamment chez les logiciens hindoux), il y a une quantité affolante de littérature sur la question du mécanisme et de l'esprit, mais il y est toujours question d'arguments en faveur du mécanisme ou en sa défaveur. Beaucoup pensent aussi que le mécanisme est par lui-même une solution au problème du corps et de l'esprit. C'est, j'espère, un apport du présent travail, de montrer que le mécanisme ne résoud pas automatiquement le problème du corps et de l'esprit. Par contre, il rend nécessaire une *reformulation du problème* sous la

La découverte décrite ici, est que, dans ce cas, c'est-à-dire avec cette *hypothèse* du mécanisme, la physique devient réductible à la psychologie *des machines*. Le "des" peut être interprété dans les sens transitif et non transitif : en clair, il s'agit de la psychologie *concernant* les machines et *inférée* ou *postulée* correctement (par définition) par les machines elles-mêmes. On pourra, avec un peu d'informatique théorique, définir cette psychologie par le sens large des discours de la machine "auto-référentiellement correcte". Cette psychologie apparaîtra non normative : nous verrons qu'elle fera de nous des êtres encore plus inconnus que nous n'aurions pu le croire. Elle constitue une sorte de vaccin contre de nombreuses formes de réductionnisme de la psychologie humaine.

La réduction de la physique à la psychologie se fait aussi bien au niveau *épistémologique* : la physique devient effectivement une *branche* de la psychologie—science du *machine-observable*—qu'au niveau *ontologique* : la matière ou l'apparence de la matière émerge de la conscience, de l'esprit ou du mental ou même, on verra, des "paris possibles" de toutes les machines, numériques possibles.

Autrement dit, ce que j'ai réussi à démontrer, semble-t-il, c'est que si on prend vraiment au sérieux l'hypothèse que nous sommes des machines numérisables, alors on est forcé de reconnaître un renversement de l'idée naturaliste ou matérialiste, assez répandue autant chez les philosophes, les physiciens et l'homme de la rue, selon laquelle la physique serait la science fondamentale à laquelle les autres sciences naturelles et humaines, au moins ontologiquement et donc en principe, devraient pouvoir se réduire. Je résume ce théorème par :

$$comp \Rightarrow renversement,$$

où *comp* désigne le "computationnalisme", un nom souvent donné au "mécanisme numérique" et *renversement* désigne le renversement psychologie/physique. Le résultat est que ce n'est pas la matière qui est primitive et la conscience qui émergerait de l'organisation de la matière, mais c'est l'inverse : la conscience serait plus primitive et c'est la matière ou plutôt l'apparence de l'organisation matérielle qui émergerait des expériences possibles des consciences possibles ; et cela dans un sens suffisamment précis que pour dériver la physique (science de la matière) de la psychologie (vue comme science très générale des expériences de la conscience, ou plus positivement, des discours stables des machines sur elles-mêmes : la physique, mais non

forme d'une nécessaire justification des croyances en l'apparence d'un monde matériel, physique ou substantiel (pour anticiper en une phrase le principal résultat du travail).

la géographie[2], appartenant nécessairement—ce qu'on va démontré ici—à ce discours autoréférent).

À ce stade, une personne qui, pour diverses raisons, serait persuadée de la véracité du matérialisme[3] contemporain, peut toujours estimer que ce travail constitue une réfutation du mécanisme. Cela pose néanmoins problème pour beaucoup parce que le mécanisme est, implicitement ou explicitement, la philosophie adoptée par la plupart des matérialistes.

En ce qui me concerne je ne dis rien. Mes *opinions* philosophiques restent et resteront privées. Je montre cependant dans la partie plus technique de la thèse que l'on peut déjà extraire assez bien de données qualitatives et quantitatives de la physique montrée dérivable de la psychologie des machines. Si on confronte alors ces résultats avec les théories de la physique usuelle, empirique et moderne—la mécanique quantique notamment—on peut y voir un début de confirmation empirique de cette psychologie, et donc du renversement.

La thèse, par l'éclaircissement apporté aux problèmes de l'interprétation des faits physiques (quantiques) conduit *de facto* à juger à la fois le renversement et sa raison logique, le mécanisme, comme étant plausibles.

Une dernière remarque s'impose concernant le rationalisme et l'interdisciplinarité.

Le travail se veut "rationaliste". J'apprécie, comme Karl Popper, opposer le rationalisme à l'élitisme. Le rationalisme est une forme d'espoir en la raison de l'autre ; c'est l'espoir que l'autre aura la politesse de vous écouter et d'accepter vos résultats ou de vous faire voir vos erreurs, ou de vous dire, au moins, que le sujet ne l'intéresse pas. Popper écrit :

> Faith[4] in reason is not only a faith in our own reason but also—

[2]La physique devenant l'étude de ce qui est en principe observable par *tout* observateur. L'existence de la lune n'étant (vraisemblablement) pas une *loi* de la physique. À ce stade on peut craindre que les lois de la physique se ramènent à des vérités triviales, nous verrons que les contraintes du mécanisme détrivialisent cette physique introspective.

[3]Le matérialisme, tout au long de ce travail, sera pris dans le sens faible de la doctrine philosophique qui postule l'existence d'un univers substanciel (fait de choses qui obéiraient à des lois indépendantes de nous).

[4]La foi en la raison n'est pas seulement la foi en sa propre raison mais aussi—et même plus—foi en la raison des autres. Ainsi un rationaliste, même s'il se croit intellectuellement supérieur aux autres, va rejeter tout argument d'autorité puisqu'il est conscient que, si son intelligence est supérieure à celle des autres (ce qui *lui* est difficile de juger), cela est ainsi seulement dans la mesure où il est capable d'apprendre à partir de ses propres erreurs et à partir des erreurs des autres, et que l'on peut apprendre dans ce sens seulement si on prend les autres et leurs arguments au sérieux. Le rationalisme est donc pieds et poings

> and even more—in that of others. Thus a rationalist, even if he believes himself to be intellectually superior to others, will reject all claims to authority since he is aware that, if his intelligence is superior to that of others (which is hard for him to judge), it is so only in so far as he is capable of learning from his own and others people's mistakes, and that one can learn in this sense only if one takes others and their arguments seriously. Rationalism is therefore bound up with the idea that the other fellow has the right to be heard, and to defend his arguments. (Karl R. Popper[5]).

En particulier je pense que la raison est une chose universelle et universellement profitable. Il n'existe pas quelque chose comme la science qui serait clairement séparé des autres activités humaines. Je crois seulement qu'il y a des gens qui ont une *attitude scientifique*, laquelle n'est qu'une forme de modestie et d'honneteté avec soi-même et avec les autres. Cette attitude ne dépend d'aucun domaine particulier. J'ai un slogan tout prêt :

Il y a des jardiniers plus scientifiques que des astronomes.

Et j'aurais pu dire astrologue à la place de jardinier[6].

Aujourd'hui il existe une sorte d'abîme maintenu artificiellement entre les sciences humaines et les sciences exactes. Par exemple, soi-disant pour combattre l'usage élitiste des mathématiques, l'idée est venue à un ministre[7] de supprimer de nombreuses heures de mathématiques dans diverses sections de l'enseignement secondaire. De même on supprime de plus en plus d'heures de mathématiques dans les sciences humaines, et on finit par décourager les professeurs d'enseigner les démonstrations—c'est-à-dire les explications—des formules du cours de mathématiques. C'est évidemment l'inverse qu'il faudrait faire, quitte à enseigner autrement, d'autres mathématiques, dans les sections humaines.

En s'interdisant l'usage généralisé de la raison, qui est toujours déductive ou interrogative, et des mathématiques, on contribue non seulement à rendre les sciences humaines moins exactes et les sciences exactes moins humaines, mais, comme cela devrait être clair d'après le présent travail, on rend surtout les sciences humaines moins humaines et les sciences exactes moins exactes.

---

liés avec l'idée que l'autre a le droit d'être entendu et de défendre ses arguments.

[5] *The Open Society and Its Enemies*. London, Hutchinson, 1950 .

[6] On consultera le beau livre de Suzanne Blackmore "In Search of the Light" pour un exemple d'une contribution scientifique, certes un peu négative, à la parapsychologie (Prometheus book, New-York, 1996).

[7] L'idée est défendue par Claude Allègre, exposé dans son livre "La défaite de Platon" (Fayard, Paris, 1995) et appliquée lorsqu'il devint ministre de l'éducation.

Notons bien que je ne prétends pas que la raison soit tout, ou qu'elle soit une sorte de panacée universelle. Je dis seulement que l'on peut considérer la raison comme le degré zéro de la politesse, celui qui permet d'évoluer et de progresser, dans la recherche de la connaissance, quitte à faire courageusement de temps à autres, des plus ou moins grands retours en arrière, comme en révisant ses croyances ou en abandonnant un préjugé.

La raison n'est pas suffisante pour le progrès de la connaissance. Il faut encore l'inspiration, l'attention, l'imagination, le courage, etc. La raison n'est pas suffisante mais elle est nécessaire pour la communication des résultats aux autres.

Concernant encore l'interdisciplinarité, j'aime souvent citer Descartes. Il a écrit :

> Il faut donc bien se convaincre que toutes les sciences sont tellement liées ensemble, qu'il est plus facile de les apprendre toutes à la fois, que d'en isoler une des autres.

J'espère que la présente contribution illustrera à quel point Descartes est inspiré sur ce point. Le travail, comme par ailleurs les travaux conjoints en mécanique quantique d'Einstein, Podolski et Rosen d'une part, et de Bell d'autre part, illustre aussi le caractère artificiel de la frontière entre la science et la philosophie, ou même entre la science et la théologie. On reviendra sur ce point.

Le tracé d'éventuelles frontières entre les sciences et les philosophies repose toujours sur des postulats philosophiques, avoués ou non.

L'avantage ici encore de vous raconter brièvement le cheminement de l'enfance de ma pensée, est que les enfants sont naturellement interdisciplinaires : ils n'ont pas encore subi le lavage de cerveau du "spécialisme académique".

Les enfants posent des questions sans se soucier de savoir où ils mettent les pieds.

Chapitre 2

Le secret de l'amibe (1961 → 1971)

que fais-je ici dans les miasmes
petipetit lilliputien
pris de terreur et parfois d'asthme
devant ces tonnes de machins ?
Gaston Compère, Géométrie de l'absence.

Ce qui va suivre constitue évidemment une vue partielle du passé. Je ne raconte pas ma vie, seulement des événements épars illustrant l'enchaînement des idées et des questions qui sont à l'origine de la découverte.

Certains pédiatres prétendent que la première *crise métaphysique* ou la première inquiétude concernant la mort, survient chez l'enfant à l'âge de 4 ans. Peut-être. Je me souviens de la terreur que m'inspirait le soir et la nuit, et j'exigeais de mes parents une sorte d'assurance que je me réveillerais le lendemain matin.

Les parents, avec le soucis bien intentionné de calmer les inquiétudes des enfants leur racontent des histoires. Etant né en Allemagne, régulièrement une nourrice me lira, en allemand, des contes germaniques, essentiellement ceux de Grimm et je ne suis pas sûr qu'ils aient appaisé mes inquiétudes.

Je devais alors avoir 5-6 ans lorsque, je m'en souviens, mon père par une sorte d'inadvertance sans doute, ou alors simplement fatigué parce que je l'assénais de questions sans jamais m'arrêter, m'apprit que Saint-Nicolas n'existait pas.

'Et le Père Noël ?' 'Non plus' me dit mon père semblant un peu attristé par mon air incrédule et attérré. Ainsi s'écroulait ma première théorie—ou ontologie, mythologie, théologie, croyance, rêve ..., appelez cela comme vous voulez, à cet âge la précision eut été prématurée.

'Et les fées ?' demandai-je. 'Non plus.' 'Mais alors, les anges, tout ça ... ?' Et là je vis à l'air de mon père que je venais encore de poser une question embarrassante. Après un long soupir il m'expliqua qu'il n'y croyait plus vraiment, aux anges et à Dieu, mais que mes cousins et mes oncles et tantes y croyaient. Cela m'étonna d'autant plus que les fées n'étaient pour moi que des anges féminins disposant d'une baguette magique. J'aimais les fées et les anges parce qu'ils volaient (si j'avais developpé cette tendance je serais devenu aviateur), mais surtout parce que les fées et les anges étaient immortels.

A présent je découvrais que les adultes pouvaient avoir des croyances différentes. Cela me choquait profondément. Et si mes cousins pouvaient croire aux anges, n'était-ce pas mon *droit* d'y croire aussi, ainsi qu'aux fées ?

Mon père m'expliqua que d'une certaine façon c'était sûrement mon droit de croire en quoi je voulais croire, mais qu'il n'était pas évident que cela soit dans mon intérêt. Si on croit à des propositions fausses on risque déceptions et déconvenues. Je trouvais pathétique l'idée de croire à du faux et cela me donnait le cafard. C'est à partir de ce moment que j'essaierai de me conformer à la règle : éviter à tout prix de croire à du faux.

Evidemment la vérité peut faire peur. En particulier l'idée que j'étais mortel me semblait être à la limite de l'acceptable. Mais l'idée de croire à du faux par peur du vrai m'inquiétait davantage. Je me promis donc de toujours chercher la vérité, aussi effrayante pût-elle être. Savoir vaut mieux.

Savoir vaut mieux : d'accord. Mais est-ce possible ? Sûrement cela n'est pas facile.

D'abord je constaterai que la nuit, pendant les rêves, j'étais capable de croire à n'importe quelle fausseté. Je souffrais par ailleurs, comme beaucoup d'enfants, de troubles du sommeil, qui avaient été confirmés par électroencéphalographie, et mes rêves étaient anormalement réalistes. Ce superréalisme était agréable lors des beaux rêves mais il devenait vraiment inquiétant lors des rêves étranges ou des cauchemars. Ce doute qui vient du rêve, et qui concerne la possibilité même de connaître la vérité, va jouer un rôle dans toute l'histoire qui nous occupe. Cela n'a rien d'original, le rôle métaphysique du rêve apparait chez les idéalistes hindoux, Platon, Descartes, Berkeley, comme je l'apprendrai plus tard.

Ensuite, il y avait ce problème de la divergence entre les opinions de mon père et de mon oncle. Avant, tout était simple : une proposition était vraie si et seulement si mon père l'assertait[1]. Mais depuis qu'il avait fait preuve de quelques secondes de doute sur l'existence des anges et qu'il m'avait appris

[1]Je m'exprime dans un langage d'adulte, à l'époque j'aurais été bien en peine de formuler cette proposition de cette façon.

que mon oncle y croyait, lui, je me demandais vraiment qui devais-je croire.

Je demanderai à mon oncle pourquoi il croyait aux anges. Il me répondra, pour autant que je puisse m'en souvenir avec précision, qu'il y croyait parce que ses parents y croyaient ainsi que ses grand-parents, etc. Je trouvais cette réponse franchement inquiétante. En effet si son ancêtre s'était trompé, cette erreur se propagerait de génération en génération. J'admirais alors mon père pour avoir remis en doute les croyances de ses parents à lui, et je déciderai de ne jamais croire à une proposition sous prétexte que cette proposition aurait été énoncée par une personne de confiance. Je découvrais là ce qu'on appelle le principe du libre examen, principe fondateur de l'université libre de Bruxelles, université où mon père termina ses études de juriste après avoir étudié chez les Jésuites. Je ne doute pas qu'il m'ait influencé.

Je demanderai à mon père pourquoi il ne croyait pas aux anges et aux fées (je m'en fichais un peu de St-Nicolas et du Père Noël qui à ma connaissance n'étaient pas immortels). Il me répondra qu'on avait déjà regardé partout et qu'on n'en avait pas rencontrés. Suivit alors un déluge de révélations : on vit sur une boule suspendue dans l'espace, on en a déjà fait le tour, etc. Il semblait ne pas y a avoir de places pour les fées et les anges.

Afin de ne pas risquer de croire à du faux, mon intérêt pour les êtres *imaginaires* glissera vers un intérêt prononcé pour les animaux, dont personne ne remet l'existence en doute. Au retour de l'Allemagne en Belgique mes parents acheteront une petite fermette à la campagne où nous allions en vacances et le week-end. Je passerai tout mon temps à observer les hirondelles, les papillons, les fourmis, etc. Lorsque je regardais un animal, par exemple, un papillon, je m'identifiais corps et âmes à ce papillon. S'il volait c'est moi qui volais, s'il butinait, c'est moi qui butinais, et c'est moi qui m'énivrais des multiples nectars des fleurs des champs.

Un jour je pointai le doigt vers un papillon blanc et m'exclamai, m'adressant à ma sœur et à mon frère : 'Regardez ! ce papillon, je le reconnais, c'est moi ; cela fait plusieurs semaines que je suis ce papillon'. Et, avec la délicatesse bien connue des frères et sœurs, ils me diront que ce n'est pas possible "parce que les papillons ne vivent qu'un jour".

Ce fut un choc. Cela me rappelait que si les hirondelles et les papillons volaient, comme les anges, il n'en étaient pas moins mortels, comme moi. Mais eux semblaient vivre nettement moins longtemps que moi, ce qui me dérangeait.

A l'époque, en effet, lorsque je m'identifiais à un animal, cette indentification opérait en temps réel : je n'imaginais pas encore que, du point de

vue du papillon, un jour pourrait paraître très long. Donc si un papillon vivait un jour, m'identifiant au papillon, je vivais un jour aussi. Et pas plus. Et ça, ce n'était vraiment pas rigolo. Je deviendrai quasi-maniaque sur l'âge maximal possible des animaux. Chaque fois qu'on me parlait d'un nouvel animal je m'enquérais de sa longévité. Je fus assez déçu de découvrir qu'en gros les grands animaux vivent plus longtemps que les petits, auquels hélas je m'identifiais davantage—j'étais moi-même assez petit, en ce temps.

C'est alors que je ferai une authentique découverte révolutionnaire. J'avais un compagnon canin qui était mon confident des inquiétudes métaphysiques et mon partenaire dans la quête du vrai. J'essayai un beau jour de lui montrer une petite araignée rouge (en fait un petit acarien de jardin), sans parvenir cependant à attirer son attention. Je concluai que l'acarien était trop petit pour que mon chien puisse le voir, et tout-à-coup, m'identifiant à mon chien, me vint à l'esprit que les fées et les anges étaient peut-être juste trop petits pour que nous puissions les apercevoir.

Je présentai aussitôt cette théorie à mon père. J'étais content, non pas d'une preuve de l'existence des fées mais d'une preuve que mon père ne pouvait pas être sûr de l'inexistence des fées et des anges. Je me faisais un peu l'avocat du diable, pas que je voulais contredire mon père à tout prix mais plutôt montrer que mes cousins et mon oncle n'étaient, peut-être, pas tout à fait dans le faux. 'Même si on a cherché les anges partout sur la terre et qu'on en a pas vu, cela ne prouve rien. Peut-être les anges et les fées sont-ils simplement trop petits pour qu'on puisse les apercevoir'. Je lui expliquai mon expérience avec mon chien. Mon père, qui décidément avait réponse à tout, m'apprit qu'on avait cherché dans cette direction aussi. Il me parla du microscope, et, ... et c'est cela qui me surprit le plus, il m'apprit qu'on avait effectivement découvert une multitude de petits animaux invisibles *à l'œil nu*. Il prit alors une feuille de papier et me dessina une amibe. Je tomberai de suite amoureux de cette adorable petite créature, multiforme et si facile à dessiner.

Et bien sûr la question fondamentale à présent, était de savoir combien de temps vit une amibe.

Avec ma croyance selon laquelle plus un animal est petit moins il vit longtemps, je ne me faisais guère d'illusion : elle ne doit pas vivre bien longtemps, la petite amibe.

A cette question du temps de vie de l'amibe, mon père, avec une sagesse infinie, se contentera de m'expliquer que l'amibe, après avoir bien mangé des animaux encore plus petits (!) pendant une journée, au lieu de mourir, bêtement, comme un papillon, se divisait en deux. Au lieu de mourir et de

disparaître, l'amibe se divise et donne naissance à deux amibes. C'est presque l'inverse de la mort. 'Mais alors, elles sont immortelles ?'

Mon père ne me répondra pas.

Je demanderai, notamment à mon grand frère et à ma grande sœur de me rapporter de l'école le plus de documents possibles sur les amibes. Ce qu'ils feront très gentiment. Je commencerai alors à écrire (c'est-à-dire, à griffoner dans tous les sens) un livre : "Le monde invisisible". L'idée était que s'il existe des mondes invisibles—et l'existence de l'amibe prouvait l'existence de tels mondes—alors on ne pouvait pas avoir de certitude concernant l'inexistence de quoi que ce soit. Mon oncle avait peut-être raison, in fine, au sujet des anges. Mais l'amibe était surtout un témoin tangible qu'il se pouvait que certains animaux soient immortels. Je passerai alors mon temps à prier mes parents pour qu'il m'offre un microscope. Et, avec ce microscope qui finira par arriver, je chercherai des amibes. Je découvrirai les euglènes et surtout les paramécies, beaucoup plus faciles à observer et je marquerai d'une pierre blanche les jours où je parvenais à observer une paramécie se divisant en deux. Et bien sûr, quand j'observais une paramécie, je devenais cette paramécie, et quand elle se divisait en deux, je me divisais en deux aussi. La question fût de savoir si la paramécie survivait à cette division.

Que se passait-il exactement ?

Ceci aboutira à ma première conférence publique sur les amibes. Bien que je sois mû par les questions que je me pose, j'ai toujours eu un enthousiasme immense pour faire des exposés oraux, des cours, des conférences, même sur des sujets éloignés de ceux qui me préoccuppent. J'avais donc déjà fait quelques exposés oraux, notamment sur les minéraux, mais, en 1963, à l'âge de 8 ans, on m'encourage de faire une conférence à l'école sur les microbes.

Intitulée "L'amibe, l'euglène et la paramécie", j'ai retrouvé dans un carnet un résumé succint :

> Mes amis je vous le dis, dans cette pièce, nous ne sommes pas 24 mais nous sommes quelques millions[2]. L'éléphant voit-il la petite araignée rouge ? Pourrait-il exister des êtres vivants si petits qu'ils nous seraient invisibles. Y aurait-il un monde invisible et un tunnel pour l'explorer ? Aussi incroyable que cela puisse paraître : oui. Le microscope est le tunnel, les microbes sont la

[2] J'aimais déjà les propositions paradoxales, c'est-à-dire les propositions vraies mais un peu étonnantes. "Nous" désignait évidemment dans mon esprit les élèves de la classe avec le professeur *et* les microbes dans la classe. Le "million" devrait être remplacé par un nombre bien plus grand, si on voulait être plus exact.

> découverte. L'amibe, l'euglène, la paramécie, la vorticelle, le stentor, la bactérie, l'ovule et le spermatozoïde, protozoaires de chez nous (!). L'alimentation, la digestion, l'excrétion, sensibilité diverse (l'œil de l'euglène), et ... la *reproduction*.
>
> Question : Combien de temps vit une amibe ? un jour ou toujours ? Si elle vit deux jours elle vit toujours. (Athénée Robert Catteau, chez le professeur Verschaeve).

Je deviendrai de plus en plus obsédé par cette question de l'immortalité de l'amibe. Je passerai, les deux années suivantes, la moitié de mon temps libre à me promener pour recueillir toutes les formes d'eaux possibles (égout, purin, mare, étang, rivière, flaques de toutes sortes) et l'autre moitié à l'observation de ces eaux au microscope. Je m'identifiais toujours complètement aux micro-organismes que j'observais, et j'essayais de *ressentir* ce qui pouvait bien se passer au moment de leur division. J'échafauderai un nombre inimaginable de théories illustrant le caractère immortel des unicellulaires sans parvenir à être convaincu par aucune d'entre elles. L'effort consistant à passer des fées aux amibes avait été mû par ma crainte de croire en l'existence de choses inexistantes, et je ne voulais à aucun prix commencer à croire que les amibes étaient immortelles alors qu'elles ne le seraient pas.

Une certitude s'était pourtant dégagée : SI une amibe vit deux jours ALORS elle vit toujours[3].

Il restait donc à montrer, pour que l'amibe soit immortelle, qu'elle survit à *une* division.

Un exemple de théorie allant dans ce sens était ce que j'appellais "le principe de l'inspecteur" :

PAS DE CADAVRE donc PAS DE MORT

Selon l'inspecteur, quand l'amibe se divise, elle ne laisse pas de cadavre, donc personne ne meurt dans la division, donc l'amibe survit. Mais le raisonnement n'est pas valide. Quand une hydre mange une amibe, elle la digère et il ne reste pas de cadavre non plus. Difficile pourtant de croire qu'elle a survécu à la digestion. Le principe de l'inspecteur s'écroulait.

Ma théorie ou mon argument de base en faveur de l'immortalité de l'amibe ou de la paramécie était directement lié à l'expérience consistant à me mettre à la place d'une paramécie concrète, à m'identifier à elle en l'observant attentivement au microscope. Évidemment, je me heurtais à une difficulté de taille. Il semble que je passe de un à deux. Comment est-ce possible ? Qui est

[3]En fait l'amibe commune se divise en moyenne toutes les 50 heures environ, mais pour simplifier je continuerai à parler comme si elle se divisait toutes les 24h.

la paramécie de départ parmi les deux nouvelles ? Les deux ? Une des deux ? Laquelle alors ?

Et surtout, si je deviens une des deux paramécies, comment pourrais-je convaincre l'autre, vu qu'elle aussi pourrait très bien prétendre être moi ?

Complètement émerveillé ici, pris dans une sorte de vertige quasi-extatique, je ressentais quelque chose d'aussi extraordinaire qu'incommunicable.

Mon sentiment était que l'amibe survivait à sa division (et donc a toutes les divisions et donc qu'elle était immortelle) mais comme elle devenait deux, chacune des deux amibes résultantes ne pouvaient convaincre l'autre qu'elle avait survécu, où "elle" désigne l'amibe de départ. D'où l'incommunicabilité.

Si une amibe ne parvient pas à convaincre une de ses congénères de sa survie ou de son immortalité, combien plus lui sera-t-il difficile de convaincre un être humain.

Et combien plus sera-t-il difficile pour moi de convaincre un être humain de l'immortalité de l'amibe, si immortalité il y a. Plus j'y réflechissais plus il me semblait que cette immortalité, si immortalité il y a, est condamnée à rester secrète. Cela expliquait à mes yeux le silence prudent de mon père.

Une confirmation étonnante viendra lorsqu'on m'offrira le très beau livre de Ames & Wyler "Les merveilles de la vie" avec une préface de Jean Rostand et de merveilleuses illustrations de Charles Harper—c'est aussi le dernier livre avec lequel je dormirai !

Ce livre contenait un chapitre entier consacré à l'amibe. J'ai d'abord cru—distrait par la somme d'informations nouvelles qu'il contenait—qu'il n'abordait pas la question de l'immortalité des protozoaires, mais un jour, je tombai sur la photographie d'une paramécie dont la légende était "La paramécie est-elle immortelle ?". Je ressentis d'abord un soulagement parce que je voyais, qu'on pouvait au moins poser la question, ensuite, un étonnement : 'voilà un livre qui répondait à énormément de questions et qui là se contentait de poser *la* question". Cet étonnement s'est mué en une confirmation que l'immortalité de la paramécie, si immortalité il y a, ne peut qu'être qu'une nécessaire interrogation : un pari au succès incommunicable.

Ames et Wyler étaient aussi prudents que mon père. Je me demandais si j'allais réussir à être aussi prudent qu'eux. Quelle malchance quand même : je découvre une vérité fondamentale et il semble qu'il soit interdit de la communiquer. Je devrai attendre 1971, pour sortir de cette impasse et commencer à mesurer la part communicable de cet incommunicable secret de l'amibe.

Remarquons bien que jusqu'à présent je ne m'étais jamais posé la question de savoir de quoi une amibe serait faite, ni de quoi moi je serais fait. Il

me semblait que la question ne dépendait pas vraiment de ce dont les choses seraient faites. Je ne m'imaginais pas moi-même comme fait de quelques choses. Le problème de l'immortalité me semblait être une question de biologie, ou de psychologie ou de théologie, pas une question de physique. Il n'en demeurait pas moins que je me demandais *comment* une amibe opérait pour se diviser en deux. De plus l'argument en faveur de l'immortalité *de* l'amibe d'une part et surtout de l'incommunicabilité de cette immortalité *par* l'amibe d'autre part, dépendait crucialement du fait qu'après la division les deux amibes résultantes étaient rigoureusement identiques, puisque c'est seulement dans ce cas que les deux amibes semblent se contredire en prétendant avoir survécu, une, chacunes !

La lecture du Ames & Wyler, accompagnée de livres de Jean Rostand que m'avait passés mon père ainsi que des excellents manuels de Jean-Pierre Vanden Eeckhoudt—professeur à l'athénée Robert Catteau—me conduira d'étonnements en étonnements. J'y apprends les mots magiques qui décrivent les phases principales de la division cellulaire : prophase, métaphase, anaphase, télophase ; ainsi que leur signification en termes de chromosomes. J'y apprends surtout que je suis moi-même constitué d'une colonie d'amibes sociales ! Notre qualité d'organisme pluricellulaire me posait problème : comment moi, société d'amibes, pourrais-je encore m'identifier à une amibe individuelle ? À moins que l'amibe elle-même ne soit à son tour une société de sous-microbes, et ainsi de suite ? Je serai donc naturellement conduit à m'intéresser à la chimie et à la physique atomique.

Je me poserai au sujet de la matière une question qui m'empêchera de vraiment prendre au sérieux l'idée d'atome pour ne pas dire l'idée même de matière. J'avais d'abord imaginé les atomes comme des sphères hyper-lisses et hyper-dures ; puis j'appris que les atomes étaient eux mêmes constitués d'électrons tournant autour de noyaux constitués de protons et de neutrons que j'imaginais à leur tour comme des sphères hyper-lisses et hyper-dures. Mais nulle part dans les livres n'apparaissaient ces sphères dures et lisses. Il semblait que l'on pouvait toujours diviser la matière et que la recherche d'une particule ultime était vaine. Mais en même temps il me semblait que si on trouvait une ultime particule, que serait-elle d'autre qu'une sphère lisse et dure à son tour, et alors de quoi cette sphère là serait-elle faite ?

L'idée même de matière me semblait complètement vide de capacité explicative. La notion de matière me semblait apporter beaucoup plus de questions que de réponses et elle semblait menacer, peut être, l'unité de l'amibe.

C'est dans le livre de Joël de Rosnay que j'apprendrai l'existence de l'ADN[4], la gigantesque molécule d'acide désoxyribonucléique qui est une longue chaine en forme de double hélice, "comme dans un certain château de la Loire", constituée de la répétition de molécules prises parmi l'ensemble {Adénine, Thymine, Cytosine, Guanine} donnant un très long mot du genre AATGGCTATGGACCTCAG.... Et c'est dans ce livre que j'apprendrai comment ce mot vu comme suite de triplets AAT GGC TAT GGA CCT CAG.... est traduit en ARN, un autre acide nucléique, lui même traduit en un "mot" : la protéine, constituée de petites molécules, les acides aminés, choisies dans l'alphabet de 20 "acides aminés". J'apprendrai comment les protéines, dont les enzymes, s'occupent de tout le reste : de la synthèse des petites molécules (acides aminés, nucléotides, sucres), jusqu'à la constitution même de la cellule.

Cela suscitait toutefois davantage de questions, comment savions-nous cela ? Qu'est-qu'une molécule ?

En fait le "Joël de Rosnay", et la revue Science & Vie, furent un tremplin pour le livre qui va devenir ma bible de base pour les années suivantes (1968 et suivantes) : l'édition française, dirigée par François Gros, préfacée par François Jacob du livre de James D. Watson "Biologie moléculaire du gène". Dans ce livre je vais pouvoir me faire un aperçu de l'incroyable danse moléculaire qui s'opère, non pas chez l'amibe, mais chez une créature plus petite : la bactérie Escherichia Coli.

Le Watson fut tellement *bible* que dans mon vocabulaire personnel le mot "Watson" est devenu synonyme de Bible. Rétrospectivement le "Ames et Wyler" fut mon premier *Watson*, mais à l'âge où je l'ai lu je crois que je ne me posais pas la question de savoir qui écrit un livre.

Le "Biologie moléculaire du gène" avait beau être *le* Watson, il fut rapidemment flanqué par un nouveau *Watson* : le "Chimie générale" de Linus Pauling, et cela augura un âpre conflit dans mon esprit qui va perdurer lui aussi les années qui vont suivre.

D'une part le Watson me donnait l'impression d'une danse formidable de molécules, danse néanmoins parfaitement codée par l'ADN et décodée par la cellule, me permettant de voir dans cette danse essentiellement une machinerie, où l'identité de chaque molécule ne contribuait en aucun point à l'identité de l'organisme entier. Ceci étant admis, cela fait sens de dire que l'amibe, qui répète de façon beaucoup plus compliquée mais similaire

[4]Bien sûr Ames & Wyler en parlent aussi, mais je ne comprenais pas vraiment de quoi il s'agissait. De plus le Ames & Wyler s'ouvrait automatiquement sur le petit chapitre consacré à l'amibe !

dans le principe la danse moléculaire de la bactérie, se reproduit elle-même *machinalement*. J'observais ainsi, à travers la génétique moléculaire, une *implémentation*[5] d'une solution au problème de savoir comment une amibe fabrique une autre amibe identique à elle-même. Concernant le problème du temps de la vie de l'amibe cela allait dans le sens de la survie de l'amibe à sa duplication (c'est vraiment une duplication fidèle), et donc vu le "théorème de base" exposé en 1963[6], elle est immortelle. Ceci est vrai indépendemment du fait qu'elle ne peut pas le communiquer, et indépendemment du fait que "je" ne peux pas non plus le communiquer.

Le Linus Pauling —le livre— était une sorte de preuve concrète qu'il ne m'était pas permis de communiquer la vérité de l'immortalité de l'amibe. Watson a dit "les cellules obéissent aux lois de la chimie". Pour être sûr du caractère vraiment "machinal", discrètement causal si vous voulez (je n'avais pas encore le concept du *digital* ou du *numérisable*) de l'autoreproduction, il fallait s'assurer du caractère discret et machinal de l'activité des molécules elles-mêmes. Or si le Linus Pauling abondait dans l'aspect discret, quantifié de la nature chimique, avec les atomes, les niveaux discret d'énergie, tout cela semblait reposer sur ... des mathématiques, et même des nombres réels, du continu, des équations différentielles. Ça alors ! De quoi s'agit-il ?

Le conflit entre le Watson et le Linus Pauling était une vraie guerre concrète qui prenait toutes ses proportions à l'approche des vacances scolaires. Comme toujours, à ma grande joie, on allait à la campagne, et la question était : est-ce que j'emporte le Watson ou le Pauling ? Je savais que si je prenais les deux je passerais les vacances à hésiter entre eux. Comme pendant l'année scolaire en ville, je passerais mon temps à feuilleter continuellement l'un et l'autre sans me décider et en restant coincer dans des abîmes de perplexités.

Cela allait au point de lire d'autres ouvrages, des "non-Watson", allant des journeaux *Tintin* et *Spirou* dont j'étais féru, des romans policiers, Freud, Young, Ionesco, Borgès, ou encore cet "Alice au Pays des Merveilles" que j'abandonnerai à mi-chemin tellement je le trouvais barbant mais auquel je reviendrai, pourtant.

Pendant tout ce temps, j'avais gardé le goût des exposés oraux. Je ferai un exposé, au cours de biologie à l'école secondaire, sur "l'opéron lactose" (les résultats en génétique bactérienne de Jacob et Monod). Je n'avais pas fini l'exposé et le professeur me laissera continuer l'heure suivante, et l'heure suiv-

[5] Une representation informatique ; on dit aussi une implantation.

[6] Une amibe vit un jour ou toujours. Dit autrement : si une amibe vit deux jours alors elle vit toujours. Ou encore : si une amibe survit à une division cellulaire, alors elle survivra toutes ses divisions cellulaire.

ante, etc. Finalement il me laissera parler pendant plusieurs semaines. Un jour le professeur résumera mon exposé qui était devenu une petite introduction à la biologie moléculaire et commettra une légère erreur. Toujours soucieux vis-à-vis du vrai, je communiquerai discrètement l'erreur à mes condisciples mais l'un deux (le traître) me poussera sur l'estrade en disant : "Monsieur, Marchal veut vous dire quelque chose". Je serai très embêté et ferai part au professeur le plus poliment possible de la présence de l'erreur dans son résumé. Il sera un temps silencieux et décidera de faire corriger l'erreur à tous les élèves. Il ne m'en tiendra jamais rigueur et nous développerons une relation pleine de respect mutuel. En fait j'ai énormément de respect pour ceux qui savent reconnaître et corriger leurs erreurs. Rappelez-vous comme j'admirais mon père pour son changement d'avis. Avec Camus, je pensais que seule la bêtise persiste toujours, peut-être.

Je m'inquiétais aussi pour le choix des études universitaires bien que cela me semblait encore très loin —j'étais cependant extrêmement impatient d'aller à l'université ne fut-ce que pour poser toutes les questions qui me réjouissaient et m'oppressaient à la fois. Ferais-je la biologie ou la chimie ? Je me posais cette question pratiquemment tous les jours.

J'aurai alors la chance de pouvoir fréquenter le laboratoire de Biologie Moléculaire de l'ULB à Rhodes-St-Genèse grâce à l'amabilité de Jean Rommelaere, dont la mère était une amie de la mienne. J'y rencontrerai Jean Brachet qui dirigeait le service de radiobiologie, et j'aurai surtout l'occasion de rencontrer et de discuter avec René Thomas qui dirigeait le service de génétique des bactéries et des virus—Escherichia Coli et les phages *lambda*. Quelle rencontre opportune ! René Thomas était un biologiste qui avait découvert la logique formelle dans le livre de Lewis Carroll "la logique sans peine". Un livre formidable dont l'édition française contient de magnifiques illustrations de Max Ernst, y compris un dessin d'une merveilleuse planaire. Le plus intéressant dans le travail de René Thomas était qu'il avait montré comment simuler des circuits logiques dans une bactérie par l'intermédiaire de gênes de contrôles du génôme de la bactérie, corroborant ainsi l'intuition, qui m'était venue de la lecture de l'article de Jacob et Monod, selon laquelle la vie était une affaire de dialogues codés. Nous nous promettrons de nous revoir—il me restait 2 ou 3 années d'enseignement secondaire. Cette rencontre décuplera mon impatience d'aller à l'université.

Je continuerai cependant à m'intéresser à la chimie et au problème de la constitution de l'amibe. Du Linus Pauling je passerai au "Watson" suivant. Un vrai petit chef d'œuvre : il s'agit du livre de Michel-Yves Bernard : "Introduction à la mécanique quantique et à la physique statistique". C'était une rare introduction à la mécanique quantique destinée aux élèves du secondaire.

Finalement je retombais sur la question "qu'est-ce que la matière ?". Les organismes sont des sociétés de cellules, les cellules sont des sociétés de molécules, les molécules semblent être des sociétés de particules élémentaires, mais la relation entre ces particules semblent nécessiter une science du continu. Mais qu'est-ce que le continu ? Et que viennent faire des mathématiques élaborées ici ?

En résumé la biologie et la génétique moléculaire présentaient des indices que nous étions des machines. Notre identité biologique me semblait définie par de l'information codée et essentiellemment indépendante des matériaux utilisés, lesquels sont continuellement remplacés. Dans ce cas l'amibe se reproduit à l'identique et est immortelle car son identité réside dans sa forme et son activité et non pas dans sa substance. La chimie et les mathématiques qui se cachaient derrière jetaient un doute sur cette conception mécaniste. Même dans la "mécanique" de Newton les objets, souvent identifiés à des "points matériels" semblent agir à distance par l'intermédiaire de champs étalés dans l'espace et décrits par une mathématique faisant intervenir le mystérieux continu. Avec la mécanique quantique, cet aspect des choses semblait poussé à l'extrême : même une particule isolée ou un atome était décrit par des fonctions qui ne s'annulaient qu'à l'infini. Il ne m'était pas évident, dans ces conditions, que l'amibe puisse se reproduire à l'identique, ni même d'ailleurs qu'elle se reproduisait. Avec la mécanique quantique et le continu, il me semblait qu'il pouvait toujours subsister un filament entre les deux amibes apparentes et qu'en réalité il n'y avait qu'une amibe qui aurait seulement fait semblant de se diviser.

En 1971, à la veille du voyage scolaire à Londres, le conflit spirituel entre la biologie et la chimie est à son apogée.

D'une part il me semblait qu'au-delà de l'immortalité de l'amibe, la puissance explicative de la génétique moléculaire résidait entièrement dans la digitalité qui permettait l'utilisation de codages et permettait une explication en terme quasi-psychologique de mémoire et de transformation de mémoire et d'interprétation de mémoire. Avec du recul, l'article de Jacob et Monod sur l'opéron Lactose—repris dans le Taylor[7]—est ma première découverte de l'explication formelle du "IF ... THEN ... ELSE" des logiciens/informaticiens.

D'autre part cette explication quasi-psychologique du fonctionnement de la cellule me semblait incomplète sans clarification de la nature de la matière.

[7]Taylor J.H., Selected Papers on Molecular Genetics, Academic Press, New-York and London, 1965.

Il ne suffit pas de dire qu'il y a des choses qui obéissent à des lois, il faut encore expliquer ce que sont ces choses, d'où viennent ces choses, pourquoi ces choses obéissent à des lois, et d'où viennent ces lois.

"Les cellules obéissent aux lois de la chimie" disait Watson. À voir. Peut-être bien que c'est la chimie qui obéis aux lois de la cellule, comme si elle était le produit d'un rêve d'amibes ...

Chapitre 3

La diagonale de Gödel (1971 → 1973)

Si DA donne AA ; et DB, BB ; et DC, CC ; que donne DD ?

En 1970 j'entre en "poésie". C'est l'avant dernière année de l'enseignement secondaire. La dernière année est la "rhétorique". Mon impatience d'aller à l'université est telle que je m'inscris aux examens du jury central avec l'idée de sauter les deux années qu'il me reste à faire dans le secondaire. Je ne poursuivrai pas cette entreprise qui était en fait paradoxale : en effet, non seulement j'hésitais toujours entre la biologie et la chimie, mais le doute s'élargissait au point que j'envisageais de faire des études de philosophie.

Je suivrai en élève libre différents cours à l'université, en séchant quelques heures de cours à l'école. Je suivrai notamment les cours de chimie passionnants de Lucia de Brouckère, avec qui j'aurai l'occasion d'un peu discuter de mes hésitations mais aussi de libre-examen. Lucia de Brouckère était une figure dominante de la laïcité et du libre examen à Bruxelles. Je continuerai aussi à me rendre au laboratoire de Biologie Moléculaire à Rhode-Saint-Genèse, bien que je n'irai plus qu'à la bibliothèque.

Agacé par cette hésitation, comme je l'ai dit plus haut, je lirai finalement toutes sortes de livres, trouvés parfois au hasard d'une promenade dans une librairie. C'est la lecture du livre de Gilles Deleuze "Logique du sens" qui m'ouvrira l'esprit à Lewis Carroll et surtout à son livre "Sylvie et Bruno", que je lirai plusieurs fois d'affilée. Je reprendrai alors "Alice au pays des Merveilles", ainsi que sa "Logique sans Peine". Il m'arrive encore à présent de prétendre que l'humour anglais n'est que de la logique classique prise au sérieux : ça ne marche jamais, ce qui explique le rire. Je commencerai alors à m'intéresser davantage à la logique et aux paradoxes de la théorie des ensembles. Je savais

par ailleurs que la chimie faisait intervenir des mathématiques élaborées, et je devais constater qu'à l'école je suivais les cours de mathématiques avec un grand plaisir.

Au voyage scolaire de l'année précédente, à Amsterdam, je me souviens n'avoir acheté que des livres (anglo-saxons) de génétique et de chimie, dont le beau livre de William Hayes sur la génétique des bactéries et de leurs virus[1], ainsi que le Taylor[2] qui contenait l'article de Jacob et Monod. J'écrirai une lettre enthousiaste à Bill Hayes qui me répondra fort sympathiquement. Cette année, en poésie donc, on allait à Londres[3]. Le conflit entre la chimie et la biologie était à son maximum et à la librairie Foyles, je fuirai cette querelle intérieure en me rendant presque exclusivement au rayon de Lewis Carroll ainsi qu'aux rayons de logique et de mathématiques.

C'est là que je trouverai le petit livre rouge de Nagel et Newman "Gödel's proof"—la preuve de Gödel. Je ne savais pas qui était Gödel et de quoi traitait sa preuve, mais en feuilletant le Nagel et Newman, je compris que le travail présentait une preuve au sujet de l'existence ou de l'inexistence d'une preuve. Cela m'intrigua. Puis je compris que cet état de fait était rendu possible par l'intermédiaire d'un codage, et la ressemblance avec le codage biologique me sauta aux yeux.

Sans trop y croire, je réalisai progressivement que l'ouvrage proposait un moyen général de construire des expressions formelles[4] capables de se référer à elles-mêmes. J'étais surpris de découvrir que ces expressions étaient parfaitement bien définies par les signes qui la représentaient, comme l'amibe me semblait l'être elle-même par les molécules et les atomes qui la constituaient.

J'avais compris comment l'amibe ou Escherichia Coli se divisait, c'est-à-dire que j'avais un modèle quasi-visuel de la reproduction à l'échelle moléculaire. Je disposais donc d'une sorte de preuve qu'une amibe pouvait se reproduire à l'identique ; mais justement ce modèle, comme je l'ai déjà dit, reposait sur la façon dont les molécules interagissaient entre elles, et, à cause de cela,

[1]The Genetics of Bacteria and their Viruses, Blackwell Scientific Publications, Oxford and Edinburgh, 1964, 2nd 1970.

[2]Taylor J.H., Selected Papers on Molecular Genetics, Academic Press, New-York and London, 1965.

[3]J'apprécierai tellement Londres et Oxford, pour leurs librairies scientifiques et Lewis Carroll, qu'à partir de cette date j'irai tous les ans en Angleterre, notamment à Oxford pour ce que j'appelerai mon pélerinage Carrolien.

[4]Suite de signes capable d'être interprétée au sein d'une théorie formelle, comme les énoncés logiques, ou capable d'être interprétée par une machine, comme les programmes. Ce n'était pas une notion très claire pour moi, à cette époque.

je n'étais pas sûr que l'amibe se reproduisait, ni vraiment par elle-même, ni vraiment à l'identique. Le génôme de l'amibe, son codage génétique, était reproduit, semble-t-il à l'identique, mais cette identité et cette capacité reposaient en toute apparence sur les lois de la chimie, laquelle semblait reposer sur le continu.

Je me demandais si c'était vraiment l'amibe—la petite unité discrète à laquelle je m'étais identifié dans ma prime enfance—qui se divisait ou si c'était l'univers—que je pensais comme un gigantesque continuum insondable—qui divisait l'amibe.

Avec le Nagel et Newman en main, je commençais à comprendre qu'il était possible d'envisager des entités autoreproductrices qui n'avaient a priori aucun lien avec la chimie ou le continu ni même, apparemment, avec l'univers des physiciens ou des chimistes. Je découvrais une nouvelle sorte d'amibe abstraite qui pourrait bien être infiniment plus facile à interroger que le petit animalcule concret des eaux avoisinantes.

Pratiquement, cette "preuve de Gödel" me sembla donner la clé de mon hésitation entre la chimie et la biologie : un triomphe à vrai dire de la biologie, encore qu'elle se soit transformée en une biologie abstraite concernant des êtres formels, dont la nature n'était, par ailleurs, pas encore très claire pour moi.

(Il y avait encore le problème de savoir si, dans ce cas, je pouvais toujours m'identifier à l'amibe, mais à ce stade j'étais tellement heureux d'avoir découvert une toute nouvelle sorte d'amibe que je relèguai cette question à plus tard).

Il y avait *autre* chose dans le Nagel et Newman. Pas tant dans la *preuve* de Gödel—où apparaissaient ces entités apparemment autoreproductrices ou autoréférentes, mais dans le résultat, dans le *théorème* de Gödel, dans son second théorème d'incomplétude, plus précisément, publié en 1931. En effet, en termes un peu crus, il semblait qu'il existait à présent des machins, disons, capable de communiquer des propositions vraies (moi qui suis amoureux du vrai ![5]), capables en plus de communiquer apparamment des propositions vraies les concernant eux-mêmes, mais, semble-t-il à cause de cela même,

[5]Ou de l'idée du vrai. Il ne faut pas croire que je prétends avoir une relation privilégiée avec la *vérité*. J'apprécie proposer des définitions poétiques de la "vérité". Par exemple : la vérité est une reine qui gagne toutes les guerres sans armée. Ou alors une déesse qu'aucun Dieu ne peut déshabiller complètement. La vérité, c'est ce que vous ne lirez jamais dans aucun journal, mais que vous devinerez en lisant deux journaux indépendants, et que vous devinerez encore mieux en en lisant trois, etc. La vérité c'est la source du doute : plus on sait, plus on sait qu'on ne sait pas, comme disent Socrate et Jean Gabin.

La vérité, ce n'est que l'espoir de la conscience.

incapables de communiquer ou de démontrer[6] certaines vérités les concernant justement eux-mêmes.

Comme l'amibe, ces machins semblaient être intrinsèquement incapables d'affirmer certaines propositions, certaines vérités, les concernant.

Quelles vérités ? La consistance de soi. Le fait qu'on ne va pas communiquer du faux.

Voilà une entité honnête, qui, parce qu'elle est honnête, est incapable de communiquer qu'elle est honnête. Et donc parmi ces machins, ceux qui communiquent qu'ils sont honnêtes sont d'office malhonnêtes. J'éprouvai de suite une attraction irrésistible pour ces machins. Je les trouvai amusants et pertinants. Cette fois-ci, il ne s'agit plus seulement de biologie abstraite, mais franchement de psychologie abstraite, et cette psychologie concernait des vérités incommunicables, semblables au secret de l'amibe. Car le plus formidable, si j'osais croire ou anticiper le Nagel et Newman, était que ces machins semblaient capables de prouver que *si* ils étaient honnêtes *alors* ils étaient incapables de communiquer qu'ils étaient honnêtes, comme l'amibe vue au microscope, qui en se divisant, me communiquait qu'elle ne pouvait pas prétendre avoir survécu. Chacune des sœurs me le disait, implicitement, en pointant un pseudopode vers l'*autre amibe*. Si l'une est une autre, elles peuvent très bien être *autres* toutes les deux.

Le théorème de Gödel et la preuve de Gödel me montraient l'existence d'entités reproductrices, des amibes abstraites, ainsi que l'existence de machins incapables de communiquer certaines vérités autoréférentielles, comme la consistance de soi. Exactement ce que je cherchais. Les machins en question étaient les théories formelles, comme l'arithmétique de Peano ou le principia mathematica de Russell et Whitehead. Plus de doute, j'allais devenir mathématicien et me spécialiser en logique mathématique.

Notons qu'à cette époque je souffrais d'un immense handicap : je n'avais pas encore entendu parler de la thèse de Church, ni de l'ordinateur. Le terme "ordinateur" évoquait en moi d'immenses frigos rigides utilisés par des banquiers. Je ne savais pas encore que Babbage un siècle plutôt avait rêvé d'une machine calculant la position des étoiles. Je n'avais pas encore entendu parler de la machine de Turing. Je ne savais pas encore, que je faisais, ou que je m'intéressais à, de l'informatique comme Jourdain à la prose. En gros la

[6]Je vais toujours utiliser le terme de communication dans le sens de l'affirmation honnête ou scientifique. J'identifie ou je modélise, ici et plus loin, ce type de communication avec la preuve formelle ou formalisable.

thèse de Church est que

Tous les machins sont des machines.

Ou encore que tout ce qui est formellement calculable (et relativement communicable) peut être calculé (et relativement communiqué) par des ordinateurs (*relativement* à une théorie formalisée).

Je ne savais pas encore que les *machines* étaient de tels *machins* (l'inverse de la thèse de Church) ni que les *machines universelles*, les ordinateurs, étaient à la fois très proches des théories formelles et très sensibles à ces entités autoreproductrices. Je ne le réaliserai que bien plus tard. (De toute façon l'informatique n'était pas encore une branche à part entière à l'université, seulement des options pour mathématiciens ou ingénieurs).

Qu'avais-je vu précisément dans le Nagel et Newman ? J'explique l'idée un peu librement. L'idée technique de base apparaît dans la ritournelle plus haut : "*Si DA donne AA, DB donne BB, DC donne CC, que donne DD* ?" La réponse, bien sûr, est que DD donne DD. Autrement dit le "duplicateur" D, qui fait de A, AA ; de B, BB, ..., appliqué à lui-même "D" donne, comme *résultat*, "DD", c-à-d l'expression décrivant D appliqué à D, c'est-à-dire l'expression de départ elle-même.

Autrement dit, si un environnement est assez riche pour supporter[7] des duplicateurs, alors il est assez riche pour supporter des autoduplicateurs, lesquels sont obtenus par les duplicateurs appliqués à eux-mêmes.

Un autre exemple. Imaginons une expression décrivant dans un certain langage formel, l'opération de substituer une inconnue X par une certaine expression formelle, par exemple $\ulcorner abc \urcorner$ dans une (autre) expression formelle $\ulcorner baX \urcorner$. Les symboles de quotation $\ulcorner$ et $\urcorner$ sont utilisés comme guillemets au sein du langage formel : ils empêchent l'évaluation de l'expression qui est quotée. On utilise implicitement de telles opérations de substitutions lorsqu'on utilise l'opération "chercher/remplacer" du traitement de texte avec un ordinateur. Dans ce langage formel cela peut s'écrire de la façon suivante :

$$subst(\ulcorner abc \urcorner, \ulcorner baX \urcorner)$$

Et considérons un *machin* capable d'interpréter cette expression formelle, c'est-à-dire d'évaluer le résultat de la substitution décrite, en l'occurence *baabc*. Il est entendu que l'opération *subst* remplace le ou les 'X' de la deuxième expression par la première expression.

[7] Si on me permet cet anglicisme. J'aurais pu utiliser le terme moderne "implanter" ou "implémenter" (autre anglicisme) ou de "véhiculer". L'idée est que l'environnement joue le rôle d'un ordinateur capable d'exécuter l'opération de duplication incarnée dans un duplicateur.

C'est bien le 'X' de l'expression quotée apparaissant à droite qui doit être remplacé par l'expression quotée apparaissant à gauche. Ainsi :

$$subst(\ulcorner aXc \urcorner, \ulcorner baX \urcorner) = baaXc$$

Pour un tel machin, vous pouvez vous convaincre que l'expression suivante

$$subst(\ulcorner subst(X, X) \urcorner, \ulcorner subst(X, X) \urcorner)$$

est autoréférérentielle. Il s'agit vraiment d'une procédure toute simple, aux conséquences invraisemblables, comme on va le voir tout de suite.

Cette procédure pour construire une expression autoréférentielle est appelée *diagonalisation*. Le terme de diagonalisation provient du fait que si $A(x, y)$ représente une matrice ou un tableau de nombre, peut-être infini, $A(x, x)$ représente alors la diagonale du tableau de nombre. La construction d'entité autoreproductible met en œuvre *deux* diagonalisations, ou une diagonalisation appliquée à elle-même. En effet on construit '$subst(x, x)$' (première diagonalisation), puis 'x' est remplacé par '$subst(x, x)$' dans '$subst(x, x)$' (deuxième diagonalisation).

Voici une généralisation utile de la technique. Imaginons qu'on veuille trouver une expression, qui au lieu de produire une version d'elle-même, produit le résultat d'une transformation T appliquée à elle-même. Dans la ritournelle il suffit de définir un nouvel opérateur D (et je le note encore D), qui cette fois-ci appliqué à A donne T appliquée à AA, ce que je note simplement $T(AA)$. Et ça quelque soit l'expression A qu'on lui présente. Dans ce cas, D appliqué à lui-même, c'est-dire DD donne $T(DD)$, et donc l'évaluation de l'expression DD donne le résultat de la transformation T appliquée à elle-même. De même, avec la substitution *subst*, il suffit de remplacer $\ulcorner subst(X, X) \urcorner$ par $T(\ulcorner (subst(X, X) \urcorner)$ pour obtenir une expression formelle :

$$subst(T(\ulcorner subst(X, X) \urcorner), T(\ulcorner subst(X, X) \urcorner))$$

dont l'interprétation donnera à nouveau T appliquée à l'expression formelle elle-même. Encore une fois, les conséquences seront invraisemblables—comme je le montre plus bas, bien que je ne prétends pas les avoir saisies clairement tout de suite avec la première lecture du Nagel et Newman. Tout comme le Joël de Rosnay fut un tremplin pour le Watson, le Nagel et Newman fut un tremplin pour le Kleene 1952 "Intoduction to Metamathematics" et le Ladrière 1957 "Les limitations internes des formalismes", qui étaient

hélas tout deux épuisés, mais que je parviendrai cependant à extraire de la Bibliothèque Nationale, véritable acte héroïque rendu possible grâce à un ami dont le père y travaillait. Cet ami, Dominique, partageait avec moi de nombreuses interrogations métaphysiques et nous étudierons ensemble le "Ladrière[8]".

Aujourd'hui, avec les ordinateurs qui pullulent littéralement, vous avez peut-être deviné qu'effectivement un ordinateur est une sorte de machin du genre décrit ici, capable de produire correctement des substitutions, par exemple, et donc susceptible d'être "infecté" par une expression autoreproductrice ou autoréférentielle.

Une conséquence apparaît alors presque immédiatement : un ordinateur ne peut pas résoudre toutes les questions. En particulier il ne peut pas résoudre à coup sûr la question de savoir si une machine arbitraire, une fois lancée dans son éxécution, va s'arrêter ou non.

En effet, si cela était possible nous disposerions d'un de ces machins, appelons-le *ARRET*? capable, appliqué à un autre machin[9] X de décider si X s'arrête ou non.

Mais nous pourrions alors construire une nouvelle expression (un nouveau machin) comme suit :

si ARRET ?(X) alors CONTINUE sinon STOP

CONTINUE est une instruction (une expression) qui lance l'ordinateur dans une boucle infinie, et *STOP* au contraire arrête l'ordinateur.

Cette expression définit une certaine transformation T, que l'on peut substituer dans l'expression autoréférentielle de la généralisation décrite plus haut. On obtient :

subst(si ARRET ?(subst(X,X)) alors CONTINUE, sinon STOP),
(si ARRET ?(subst(X,X)) alors CONTINUE, sinon STOP)

L'évaluation de cette expression sera difficile à lire, mais pour l'ordinateur, elle sera équivalente à p avec :

p = si ARRET ?(p) alors CONTINUE, sinon STOP,

ou encore

si ARRET ?(MOI) alors CONTINUE, sinon STOP,

laquelle est capable de décider si elle s'arrête (quand on la donne à l'ordinateur) et dans ce cas de continuer, ou de décider qu'elle ne s'arrête pas, et

[8]Ladrière, J., 1957, Les limitations internes des formalismes, E. Nauwelaerts, Louvain, et Gauthier-Villars, Paris.

[9]Ou à la description formelle de cet autre machin.

dans ce cas de s'arrêter. Cela est absurde, et donc il n'y a pas de machin du genre *ARRET*? possible. En terme de machine : aucune machine n'est capable de décider d'une façon générale, lorsqu'on lui présente une (description d'une) machine, si celle-ci va s'arrêter ou non.

Gödel montre de façon similaire que les théories formelles assez riches (en terme de proprositions arithmétiques *mécaniquement* prouvables) sont capables, pour tout prédicat[10] $P(x)$ il existe une proposition précise q telle que la théorie formelle peut démontrer $q \leftrightarrow P(\ulcorner q \urcorner)$. La proposition q est autoréférentielle—elle se réfère à elle-même. Il s'agit là d'une forme élémentaire de ce qu'on appelle *le lemme de diagonalisation* dans la littérature[11]. La preuve de ce lemme nécessite seulement de démontrer que la théorie est capable de prouver des vérités élémentaires concernant la substitution. Cela explique la portée gigantesque de ce lemme *de l'autoréférence*.

On obtient par exemple à peu de frais le théorème de Tarski—la non définissabilité de la vérité—en traduisant dans le langage de la théorie la proposition paradoxale d'Epiménide :

Je ne suis pas une proposition vraie

On ne peut plus, à l'instar du PRINCIPIA MATHEMATICA de Russell et Whithehead, supprimer le paradoxe en interdisant l'autoréférence, puisque celle-ci est incontournable pour les systèmes capables de manipulations élémentaires comme la substitution. Ce que Tarski prouve ainsi est que la notion de vérité (d'une proposition) pour un système formel (assez riche et consistant[12]) n'est pas définissable *dans* le système formel[13].

Par contre Gödel a montré que la prouvabilité par un système formel (assez riche) *est* représentable *dans* le système formel (ce que l'on conçoit

[10]Un prédicat est l'équivalent formel d'un adjectif définissable dans le langage de la théorie.

[11]Le terme "lemme" est utilisé par les mathématiciens pour désigner un résultat préliminaire.

[12]Un système formel ou une théorie, ou une machine générant des propositions est dite consistant(e) lorsqu'il ou elle ne prouve pas de propositions fausses ou des propositions contradictoires comme $p \,\&\, \neg p$. "$\neg p$" est mis pour la négation de p. Si p est vraie $\neg p$ est fausse, et si p est fausse $\neg p$ est vraie.

[13]De façon précise, appelons un prédicat $V(x)$ prédicat de vérité si le machin (la machine, la théorie) prouve $p \leftrightarrow V(\ulcorner p \urcorner)$, et ça quelle que soit la proposition p. Si $V(x)$ était définissable dans le langage de la théorie, on pourrait définir un prédicat de fausseté $F(x)$, ($F(x)$ est définit par $\neg V(x)$), pour lequel la théorie prouverait $\neg p \leftrightarrow F(\ulcorner p \urcorner)$ quelle que soit la proposition p. Mais en appliquant le lemme de diagonalisation au prédicat F, on exhibe une proposition q telle que la théorie prouve $q \leftrightarrow F(\ulcorner p \urcorner)$. La machine (la théorie) prouve alors la proposition fausse $q \leftrightarrow \neg q$. $V(x)$ n'est donc pas définissable dans le langage de la machine. Ce théorème, de Tarski, joue un rôle au chapitre 8.

aisément vu que la notion de preuve formelle, à la différence de la notion de vérité (même formelle), est une notion essentiellement combinatoire. Le paradoxe d'Épiménide, avec 'prouvable' à la place de 'vrai' conduit alors au théorème d'incomplétude de Gödel 1931.

En effet la proposition :

Je ne suis pas prouvable par la théorie T

est représentable dans la théorie T, supposée consistante, et est donc d'office vraie et non prouvable par T. En effet, si la proposition était fausse, vu ce qu'elle dit au sujet d'elle-même, elle serait prouvable et T prouverait une proposition fausse et donc serait inconsistante.

Je penserai alors que le théorème de Gödel illustre l'existence d'une mathématique permettant

1. de révéler le secret de l'amibe sans tomber dans le piège de communiquer de l'incommunicable. L'idée est de creuser l'analogie entre "je suis vivant", "j'ai survécu", "je suis conscient" et "je suis consistant". L'expérience par la pensée de l'autoduplication a déjà illustré la non communicabilité de la survie. Pour l'amibe, comme pour les *machins* du Nagel et Newman, il semble qu'il y ait du vrai qui soit non communicable.
2. d'offrir un cadre rigoureux où l'on peut opérer le renversement épistémologique entre la biologie (ou psychologie, théologie) et la chimie (ou la physique). Les machins en question semblent fournir une biologie et une psychologie mathématique, fondamentale, et indépendante de la chimie. Le reste du travail illustrera l'usage de cette mathématique.

Dans mon esprit, Gödel donnait ainsi raison à la biologie contre la chimie. La logique mathématique me donnait le sentiment qu'on pouvait étudier d'une façon générale les discours des machines (les machins, à l'époque !) obtenus lorsqu'elles s'observent elles-mêmes. Il me semblait que le travail des physiciens était un cas particulier que l'on devrait pouvoir justifier à partir de cette théorie plus générale. Watson disait que la cellule obéit aux lois de la chimie. Avec le théorème d'incomplétude de Gödel j'entrevoyais un modèle communicable de réalité où au contraire, c'est la chimie qui obéit aux lois de la cellule, où la cellule devient l'amibe abstraite, la petite entité autoréférentiellement correcte relativement à *un*[14] (à l'époque) environnement

[14]Il faudra attendre 1987 pour que ce point devienne tout à fait clair. L'autoréférence correcte ne sera plus définie relativement à *un* environnement universel (machine universelle) mais relativement à un environnement universel *le plus probable* ou le plus *crédible*.

universel (dans le sens de Church ou de Turing). Le théorème de Gödel accentuait largement le caractère *tangible* de la réalité mathématique, et il me semblait que la chimie pouvait être avantageusement considérée comme le produit des rêves et des discours cohérents d'amibes immortelles.

Un événement particulier est significatif à cet égard. Il s'agit d'une discussion assez animée avec mon ami Dominique sur le statut fondamental des différentes sciences. À cette époque Dominique affirmait que la physique était la science fondamentale. La discussion était âpre car il s'agissait de se décider pour le choix des études universitaires.

Selon moi la physique ne pouvait pas être la science fondamentale. L'idée était qu'on comprend davantage si on comprend comment un cerveau "regardant" l'univers produit une théorie de l'univers, qu'en comprenant la théorie de l'univers". Et si le cerveau est semblable à l'amibe ou à une théorie formelle, ce processus de compréhension ne dépend pas de la nature matérielle dont il serait constitué. De façon ultime il faudrait plutôt expliquer d'où viennent chez les systèmes formels ou les "machins", les croyances qu'il existe un univers ou qu'il existe de la matière. Sans doute aussi à cause des rêves réalistes de mon enfance, mais aussi de ma crainte de croire à l'existence de choses inexistantes, je n'ai *jamais* pris l'existence de la matière pour une vérité établie. Maintenant, avec l'apparition d'une biologie et d'une psychologie indépendante des lois de la matière, je commencais à penser que ce concept de matière devait être expliqué en termes plus primitifs de croyances chez certains "machins".

Chapitre 4

Plus noir que vous ne pensez [I] (1973 → 1977)

Il est possible de détruire quelqu'un
juste avec des mots, des regards, des sous-entendus :
cela se nomme violence perverse ou harcèlement moral.
Marie-France Hirigoyen .

Je m'inscris donc, enfin, en candidature en sciences mathématiques à l'Université Libre de Bruxelles, avec mes deux valises de biologie et de chimie, et la "preuve de Gödel" comme espoir de parvenir à relier l'un à l'autre, peut-être pas dans le sens affirmé par Watson.

Après la première heure du cours de logique je demandai au professeur de logique, X, pour ne pas le nommer, s'il comptait aborder le théorème de Gödel cette année ou plus tard parce que ... Je m'apprétais à lui faire part naïvement de ma motivation pour le théorème de Gödel, lui dire que c'est ce théorème qui m'a décidé à faire les mathématiques en tranchant le nœud Gordien du doute entre la biologie et la chimie etc. Mais je n'aurai pas le temps (je n'aurai *jamais* le temps) de terminer cette phrase. Dès qu'il entendit le nom de Gödel il m'interrompit aussitôt par "Oubliez le théorème de Gödel, il n'y a rien d'intéressant là-dessus, c'est une histoire terminée".

C'était évidemment faux. Mais à cette époque je ne le savais pas. C'est vrai que mon Kleene[1] datait de 1952, et le Ladrière de 1957. Je le prendrai au mot et impatient de me mettre à la page je suivrai dès la première année, avec son accord, accompagné de mon ami Dominique—qui s'était inscrit en physique, finalement—l'entièreté des cours de logique donnés par ce pro-

[1]Kleene S. C., 1952, Introduction to Metamathematics, North-Holland.

fesseur en première et deuxième années, quitte à brosser[2] d'autres cours. Et une certaine amitié se mettra en place. Son cours de théorie des modèles était intéressant.

Il nous conduira avec Dominique au séminaire de logique mathématique de Louvain ; mais non au séminaire de logique philosophique, bien qu'à Louvain ces séminaires regroupaient les logiciens philosophes et les logiciens mathématiciens—la logique philosophique est tout autant mathématique que la logique mathématique pourtant.

'Et la logique intuitionniste ?' lui demandai-je un jour. 'C'est tout à fait idiot' me répondit-il. 'Et la logique modale ?', 'laissez ça aux philosophes'. Etc. Le pire est que je prendrai cette attitude comme une forme d'humour et que je continuerai à l'admirer sans que je ne parvienne vraiment à comprendre pourquoi, à part que j'avais envie d'admirer un logicien qui m'apprenait toute sortes de choses intéressantes en logique, et qui faisait preuve d'une certaine forme d'humour décapant que j'appréciais.

J'arrivai ainsi sans problème en seconde licence[3](quatrième et dernière année). Entretemps je fréquentais à nouveau le laboratoire de biologie moléculaire de l'ULB à Rhodes-St-Genèse, et, bien que je passais plus de temps à la bibliothèque que dans les laboratoires, je discutais à nouveau souvent avec René Thomas et ses élèves, dont Jean Richelle dont j'étudierai en profondeur le mémoire de fin d'études, sur la "logique" du phage lambda[4].

Ladrière se souviendra de moi et m'offrira un exemplaire de son formidable livre sur le théorème de Gödel, et m'invitera aussi à Louvain pour exposer les travaux logiques et biologiques de René Thomas. Je ferai encore, de mon initiative, avec l'encouragemment de mes condisciples, réinspiré par le 'Ladrière', quelques exposés sur le théorème de Gödel, le soir à l'ULB.

Par prudence je n'avais rien promis à René Thomas, mais je finirai par demander à X, sans trop y croire cependant, s'il était envisageable qu'il dirige mon travail de fin d'étude avec la collaboration de Thomas (un travail interdisciplinaire en somme).

Je n'avais pas osé lui dire que je pensais que Thomas me proposerait d'approfondir la relation entre les systèmes décrits par des équations logiques

[2]Sècher, en belge.

[3]Si on fait abstraction que j'ai oublié de me rendre à l'examen de calcul des probabilités deux fois d'affilée et autres anecdotes de ce genre.

[4]Contribution à l'étude théorique de certains aspects de la régulation de l'immunité du bactériophage tempéré λ. Mémoire, ULB, 1974.

discrètes et des systèmes semblables décrits par des équations différentielles. Je n'avais pas non plus osé lui dire que je ruminais un projet personnel que je souhaitais proposer à René Thomas. Je voulais voir si les équations logiques de Thomas, celles qu'il parvenait à faire éxécuter par la bactérie Escherichia Coli, étaient assez riches pour calculer les fonctions récursives. En terme moderne cela reviendrait à voir une bactérie comme s'il s'agissait d'un (petit) ordinateur. Auquel cas, vous avez deviné, le théorème de Gödel s'appliquerait aux bactéries, aux cellules, et donc à l'amibe, etc.

J'hésitais car j'anticipais deux épreuves éprouvantes. Si j'optais pour le sujet suggeré par Thomas, je craignais de devoir m'affronter au monde des équations différentielles. Celui-ci me plonge usuellement dans un abîme de perplexité qui me ramène toujours aux questions qu'est-ce qu'un nombre réel, qu'est-ce que le continu, et quid du paradis de Cantor ? L'autre option (secrètement reliée au secret de l'amibe) risquait d'être un test pour savoir si les remarques de X sur Gödel étaient vraiment humoristiques ou non.

J'en étais là à cogiter, quand, sans plus d'explication, X me donne son accord pour un travail de fin d'étude en collaboration avec René Thomas. Le bougre ! Je savais qu'il était ouvert d'esprit ! Et me voilà donc face au sempiternel dilemme, l'amibe discrète ou le continu ?

Je n'ai pas dû réfléchir longtemps parce qu'aussitôt après m'être inscrit au travail de fin d'étude chez X, celui-ci me fit savoir qu'il avait changé d'avis. 'Forcing[5] ou Ensemble Admissibles ?' me propose-t-il. Et me voilà chargé de résumer un article et de lui montrer chaque semaine l'état d'avancement du travail.

Ce fut un calvaire épouvantable, qui m'a semblé durer plus longtemps que toutes les années précédentes. X passa son temps à me démontrer qu'il était aussi malin que moi idiot, en rabotant notamment toute originalité que j'aurais pu glisser, si bien que mon travail de fin d'étude n'est rien d'autre qu'un résumé *par* X, d'un article choisi dans la littérature, à l'exception d'une petite section originale de ma part, purement mathématique et technique je précise, que je parviendrai à préserver tant bien que mal.

Sans m'injurier explicitement, il parvint apparemment, au bout de ce calvaire, à me convaincre que j'étais *vraiment* "tout à fait idiot", absolument inapte à une quelconque carrière académique et il nota "mon" travail de fin

[5]Il s'agit du nom d'une technique inventée par Paul Cohen, un élève de Gödel, pour démontrer l'indépendance de certaines formules de la théorie des ensembles. Les "ensembles admissibles ont été introduits par Kripke et Platek, et sortent du cadre de ce travail.

d'étude 15, ce qui effectivement me fera refuser les diverses bourse nationales de recherche. Pour comprendre ce "15", je demanderai à X s'il avait trouvé des erreurs dans la partie originale. X, après avoir pris un air étonné, me fit savoir que de toute façon la note n'était plus changeable.

Quand à une bourse internationale elle nécessitait une lettre de recommandations. Je ne pouvais décemment pas en demander une à Thomas, que je n'ai pas osé revoir, parce que j'avais honte et que je n'y croyais plus.

L'amibe était lointaine et la logique me donnait la nausée.

Cela s'est passé en 1977. Ce n'est que récemment (2000) que j'ai compris que j'avais eu affaire à ce qu'on appelle aujourd'hui un pervers moral, ou encore un harcèleur psychologique. On les assimile parfois aux vampires et il est vrai qu'ils prennent quelque chose de votre vie. Je ne sais pas ce que j'aurais fait si j'avais compris plus tôt ce qui m'arrivait. Me plaindre ? On aurait tôt fait de me considérer paranoïaque—il y a en outre tellement d'étudiants qui se plaignent d'être mal notés. Je ne me plaindrai pas, je n'y penserai même pas : je me sentirai au contraire coupable comme si c'était déplacé que je puisse encore m'intéresser au théorème de Gödel, moi le très *certainement* idiot[6].

Complètement démoli, et intellectuellement perverti, je m'estimais presque heureux de mon sort : m'ayant convaincu que je ne pouvais pas me mesurer à des "vrais" logiciens, et ne tenant par ailleurs pas à me mesurer avec qui que ce soit, X me fera presque prendre comme bonnes nouvelles les échecs de mes tentatives d'effectuer une carrière académique.

[6]Une certitude que je parviendrai à mettre en *doute*, fort heureusement. Mais il y faudra du temps et de la chance, comme l'illustre la suite de l'histoire. Notez qu'il ne s'agit là que d'un *doute supplémentaire*, mais c'est un heureux doute que je souhaite à tout le monde.

Chapitre 5

Liberté chérie (1977 → 1987)

Si vous êtes pressé, prenez un détour.
Proverbe Taoïste

"Adieu veaux, vaches, cochons, ..." L'idée de faire une carrière universitaire avait été un beau rêve. À présent je devais gagner ma vie, et je me sentais libre et heureux.

D'ailleurs n'avais-je pas toutes les raisons d'être heureux ?
Heureux, d'abord, que ce calvaire du travail de fin d'études soit terminé. Je m'étais conditionné à faire une thèse "normale" de mathématiques pures, mais l'idée de prolonger quoi que ce soit d'académique évoquait un calvaire. Un calvaire *annoncé* vu que je l'attribuais—plus ou moins consciemment à cette époque—à mon incompétence.

Bien que je n'aie pipé mot de mes investigations pré-universitaires pendant mes études, autres que "Gödel ?", l'avalanche de "coups" que cette seule question apporta, non seulement me dégoûta de la logique, mais tout autant de la biologie. Et l'amibe gödélienne fut rangée dans le placard de mes fantasmes infantiles. Mais si le biologiste Gödelien était mort, la chimie se réveilla en moi, avec, certes, une légère appréhension des mathématiques (un comble pour un mathématicien), mais soulagé et heureux en somme qu'elle ne soit qu'un hobby du professeur de mathématique que j'allais devenir, pour les six années suivantes, et d'autres encore, dans diverses écoles et instituts de la ville de Bruxelles.

Content, donc, de retrouver ma liberté de pensée, préalable indispensable à la recherche fondamentale sérieuse.

Encore heureux, finalement, car j'avais conservé le goût pour l'exposé oral, et j'appréciais beaucoup et j'apprécie encore le métier d'enseignant, particulièrement des mathématiques.

Heureux peut-être mais malheureux sûrement, et une légère dépression va quand même lentement s'installer.

Le reste de l'histoire "intellectuelle" est un peu tortueux. Je vais essayer de résumer quelques événements principaux et je reviendrai ensuite sur le détour de la chimie quantique.

Le réveil du chimiste, en effet, me reconduira rapidemment à la mécanique quantique et cette fois-ci je prendrai conscience des bizarreries proprement[1] quantiques : l'indéterminisme, l'inséparabilité, le problème de la mesure, etc. En 1978 j'écris une "mini-thèse" sur les inégalités de Bell, dans laquelle je pose surtout des questions. Puis je suivrai un chemin probablement assez "tarte à la crême". La nature troublante de la réalité qu'illustrait la mécanique quantique, me fit chercher d'autres conceptions du monde, si possible éloignées de l'aristotélisme ambiant. Motivé par les philosophes taoistes—Lao-Tseu, Lie-Tseu, Tchouang-Tseu—j'ai commencé à suivre des cours de chinois classique . J'ai lu de même les doctrines immatérialistes et matérialistes Hindoues, Platon, etc.

Viendra alors une période assez intense où s'affronteront en moi deux conceptions de la réalité. En fait la guerre entre le biologiste mécaniste, qui n'était pas encore tout à fait immatérialiste—et l'était probablement beaucoup moins que dans mon enfance—et le chimiste mystique, ultra-matérialiste, prendra la forme d'une confrontation entre deux interprétations de la mécanique quantique, celle de Wigner, qui est a priori non mécaniste et quasi-idéaliste et où la conscience construit d'une certaine façon la réalité, et celle plus mécaniste (et a priori plus matérialiste) d'Everett où chaque conscience possible est portée par une réalité relative possible. Je reviendrai là-dessus. C'est dans la perspective plus mécaniste d'Everett que je reviendrai à l'argument du translateur (téléportation classique[2]). Il s'agit d'un retour implicite

[1] À la différence de la bizarrerie de base de toute la physique qu'était pour moi la présence du continu.

[2] Voir le chapitre suivant ou la thèse. Le télétransport (ou la téléportation) revient à se faire scanner à un niveau *suffisamment* fin de description, à se faire annihiler, et à se faire reconstituer ailleurs à partir de l'information qui a été scannée. Croire que l'amibe survit à la duplication, croire que l'on puisse survivre à la téléportation, croire que l'on puisse survivre avec un cerveau ou un corps artificiel, sont autant de façon de décrire ce qu'est, pratiquemment, la croyance au computationnalisme. Ne confondez pas la téléportation classique décrite ici et la téléportation quantique. Il y a des relations entre les deux, mais elles dépassent le cadre du présent travail. J'avais imaginé cette notion de façon indépendante et j'utilisais les termes de translation et de translateur.

à l'immatérialité de l'amibe. Je décrirai alors dans mon carnet de 1980 deux "expériences" possibles :

1. *La petite réalisation*. En gros, la compréhension de sa propre immatérialité. Dans mon carnet de 1980, je décris cela comme une expérience possible, quasi-mystique, que l'on peut faire, mais j'insiste sur le fait que l'on peut la déduire de l'argument du translateur. On peut expliquer cette immatérialité à quelqu'un qui veut bien accepter le télétransport classique comme moyen de locomotion. Et l'on n'a pas besoin de comprendre la mécanique quantique pour comprendre l'argument. Bien évidemment cela est relié à l'argument selon lequel si une amibe survit à une réplication son identité est dans sa forme et non dans sa matière.
2. *La grande réalisation*. En gros je la décris de beaucoup de façons dans le carnet de 1980, parfois en termes très "mystiques", comme la disparition de soi ou de l'univers, mais le plus souvent par l'"équation" WIGNER = EVERETT. Je reviens sur ce sujet un peu plus loin. Cette expérience est décrite comme étant exclusivement mystique et non communicable. Il s'agissait encore et toujours de la redoutable faiblesse de mes approches fondamentales et cela me désespérait. Je réaliserai plus tard que ce type d'expériences "mystiques" était un véritables leurre créé par mon esprit pour me faire admettre l'idée de l'existence d'une matière, que ma "rigueur de l'enfance" avait pourtant sérieusement ébranlée.

J'avais profondément "régressé" par rapport à mes intuitions de 1963 et 1971. Le chimiste (en moi) était matérialiste.

Il y avait quand même un progrès pédagogique. Le translateur illustrait le caractère communicable de sa propre immatérialité pour ceux qui acceptent de l'utiliser comme moyen de locomotion. Une idée que je transmettais sans doute mal avec l'amibe dans mon enfance.

Bien sûr je ne parlais jamais de ça à l'université. J'ai néanmoins testé l'argument du translateur pendant un an (1980) sur mes amis de la cafétaria Beppino—au point de les agacer parfois ! Ces conversations ainsi que des sondages d'opinion sur la question "montez-vous dans le translateur" sont mentionnés et détaillés dans mon carnet de 1980.

En ce qui concerne la grande réalisation, je me taisais. J'avais de nouveau le sentiment d'une vérité profonde mais totalement incommunicable. Je me demandais aussi à quoi servirait, dans un monde où on a faim, de découvrir un poisson si gros que personne ne pourrait le pêcher. Le problème est que je refoulais toujours l'idée de revenir à Gödel, *comme* moyen de communiquer "l'incommunicable". La situation était d'autant plus compliquée que je pensais à l'époque que la *grande réalisation*, que je reliais à une forme de mysti-

cisme matérialiste quantique, était en contradiction avec la *petite réalisation*. Je liais sans m'en rendre compte (car la justification Gödélienne de 1971 était refoulée) le mécanisme avec la matière (comme tout le monde), tout en sachant que le mécanisme entraînait un immatérialisme non délimitable (sans doute que l'"amibe" avait été moins refoulée que "Gödel").

Le progrès futur consistera, en revenant aux intuitions de mon enfance (1963 et 1971) à déplacer l'immatérialisme du coté du mécanisme et d'Everett, et le matérialisme du coté de Wigner. Mais je m'étais considérablement éloigné de cette possibilité de mouvement.

Et donc une certaine dépression s'installe après cette période intense. J'en viens à la cuisine végétarienne et j'ai de plus en plus fréquemment recours à la méditation zen pour finir par m'*immobiliser* quasi complètement fin 1980.

Bon. Je ne retrouve pas mon carnet de 1981. Ni 82, ni 83. Il y a quelques événements significatifs cependant—significatifs pour le développement de la future"thèse", j'entends.

- Mon ami André s'achète un ordinateur TRS 80 et me propose une démonstration[3]. C'est à ce moment que j'ai le TILT de la thèse de Church et que je réalise l'importance de la machine universelle. J'achète à mon tour un TRS 80 et j'étudie son fonctionnement, et l'informatique de plus en plus théorique. J'ai encore la nausée de la logique, ce qui dans le monde de l'informatique théorique est un fameux handicap. Je me rappelle rétrospectivement que X n'avait jamais mentionné la thèse de Church dans son cours de calculabilité. Une omission qui confine artificiellement le sujet et m'empêchera de réaliser la portée du théorème de Gödel dans le monde des machines digitales. Son cours ne mentionnait pas plus Church que Gödel !
- La parution du remarquable livre de Judson Webb "Mechanism, Mentalism and Metamathematics[4]" qui montre notamment en quoi le théorème de Gödel est une confirmation de la thèse de Church. Il développe "mon" intuition de 1971, mais cette fois-ci en relation claire avec la machine universelle. Je m'en rendrai compte plus tard ; je n'ouvre pas tout de suite le livre en effet, à cause de la crainte de découvrir que X s'était trompé (ou m'avait trompé) et de découvrir que le théorème de Gödel était bien vivant autant chez les logiciens mathématiciens qu'en philosophie des sciences. Je n'ouvre pas le livre à cause de la nausée que m'inspire toujours la logique mathématique.

[3]TRS = Tandy Radio Shack.

[4]Webb J. C., 1980, Mechanism, Mentalism and Metamathematics : An essay on Finitism, D. Reidel Pub. Company, Dordrecht, Holland.

- Mon amie Corinne revient des USA avec un exemplaire du livre de Hofstadter "Gödel, Escher, Bach" qu'elle me propose de lire aussitôt. En réalité, elle me propose de faire une vidéo sur le thème de l'auto-observation pour une exposition d'art—ce que nous réaliserons effectivement. Dans une sorte de moment de crise je lirai le livre d'Hofstadter, à la campagne, trois fois d'affilée. Mon sentiment premier est qu'il s'agit d'un fort beau livre et d'une originale introduction au théorème de Gödel. Il n'approfondit pas l'idée cependant, et, à la différence de Webb mais comme moi depuis 1971, il ignore la machine universelle, malgré un chapitre sur la thèse de Church[5]. En fait la machine universelle est la grande oubliée dans le travail de Hofstadter. Néanmoins j'ai trouvé pertinente son utilisation du théorème de Gödel en faveur de la possibilité de l'Intelligence Artificielle, ainsi que sa critique de l'argumentation de Lucas[6]. D'une certaine façon ce livre m'encouragera dans l'étude de l'Intelligence Artificielle (IA). Avec le temps je pense que cet ouvrage a peut-être écarté des chercheurs en IA du théorème de Gödel. Hofstadter tourne autour du pot. Du *bon* pot, mais il tourne trop vite autour et, par une sorte d'effet centrifuge, il s'écarte, avec son lecteur, de l'idée que le théorème de Gödel est vraiment important pour les sciences cognitives ; qu'il pourrait, par exemple, être le premier théorème de psychologie exacte[7], idée déjà énoncée par John Myhill dans les années 1950, comme je le découvrirai plus tard.
- La parution du Hofstadter et Dennett : "Minds'I". Dans un premier temps le livre me décourage car j'y lis ce que j'aurais pu écrire de mieux au sujet de la "petite réalisation". Malgré cela le livre deviendra manifestement un "Watson" et reste certainement la meilleure introduction à la présente thèse. Je le recommande à celui qui veut étudier mes travaux, comme je lui recommande aussi le remarquable petit livre de science-fiction "Simulacron 3" de Daniel Galouye. Mais ni Hofstadter, ni Dennett ne font le lien entre l'improuvabilité à la Gödel et l'incom-

[5]C'est elle qui permet de supprimer le qualificatif "de Turing" derrière "machine universelle". Voir le chapitre suivant.

[6]Lucas propose en 1959, un argument selon lequel le théorème de Gödel montrerait que nous ne sommes pas des machines. En fait l'argument se trouve déjà dans les notes d'Emil Post de 1921. Voir la thèse pour plus de renseignements. Voir le rapport technique IRIDIA 1995 pour une description détaillée des relations entre le mécanisme et les théorèmes d'incomplétude de Gödel.

[7]Mais Hofstadter met surtout le théorème de Gödel à la mode, et les gens "sérieux" affectent de mépriser la mode. Par exemple John Haugeland, (*Artificial Intelligence, The Very Idea, MIT Press.*) fait part de ses regrets de ne pas pouvoir aborder tels ou tels sujets, mais, sans la moindre justification, il explicite qu'il n'a aucun regret de ne *pas* parler du théorème de Gödel !

municabilité de la survie avec l'usage du translateur, au point que je continue à douter de la pertinence de l'association entrevue en 1971.

- La découverte du cannabis. En 1980, suite à la lecture d'Alan Watts, je décide de tester le cannabis sur moi. Mes parents m'ayant appris à me méfier des choses inconnues, je me contenterai de planter trois graines et pendant que les plantes se développeront, je lirai d'abord le maximum de littérature sur le sujet, aussi bien des "pro" comme Solomon Snyder que des "anti" comme Gabriel Nahas. C'est d'ailleurs la véhémence systématique de ce dernier qui me donnera le plus d'indices de l'innocuité de la substance. Cette relative innocuité, comparée au tabac ou à l'alcool par exemple, est aujourd'hui reconnue par les experts officiels de la santé de la plupart des pays Européens[8]. Il est reconnu par ailleurs que le cannabis rend supportable certaines nausées, comme celles provoquées par le traitement chimiothérapeutique. En ce qui me concerne le cannabis me rendra supportable la nausée et le dégoût que j'avais de la logique. Peut-être même l'herbe fonctionne-t-elle en entraînant une amnésie partielle rendant possible une mise à l'écart des connotations que l'on associe parfois au cours de la vie. Utile lorsque la connotation est négative. Le cannabis me permettra aussi d'écourter les scéances de méditation, pour atteindre un semblable niveau de relaxation (disons) dans un premier temps et de les éliminer dans un second temps ou, en tout cas, de les rendre très rares ; ce qui représente un gain de temps considérable, mais aussi soulage les genoux.

Avec tout ça le biologiste gödélien "ressucita" donc, d'une certaine façon.

Au moment où je me passionnais le plus pour l'intelligence artificielle je suis appellé par le service militaire. J'opte alors pour un service d'objecteur de conscience et Georges Papy, professeur au département de mathématique à l'Université Libre de Bruxelles, me propose de faire ce service au département d'algèbre (1982-84). La mission est formidable : pédagogie de l'informatique. Je donnerai des cours d'informatique à toutes sortes de publics : enfants, enfants handicapés, instituteurs, étudiants, professeurs, etc.

C'est en 1982 que je serai invité, grâce au professeur Papy, dans le cadre de mon service d'objecteur, à Arlon, en Belgique, pour exposer un moyen pédagogique d'enseigner le fonctionnement de l'ordinateur à de jeunes personnes. Je ferai un exposé de huit heures, traduit simultanément en italien : "L'ordinateur est un graphe" qui donnera lieu à une publication italienne[9] en

[8]Voir par exemple le Bernard Roques, La dangerosité des drogues, Rapport au secrétariat d'État à la Santé, Éditions Odile Jacob, 1999.

[9]L'elaboratore e'un grafo, L'insegnamento della matematica e delle scienze integrate,

1983. Cet article est l'ancêtre du "paradoxe du graphe filmé", que je décris dans un cahier de 1984, et qui constituait pour moi une façon d'illustrer la difficulté du problème de la relation matière/esprit avec la thèse du mécanisme. Le paradoxe du graphe filmé deviendra "l'argument" du graphe filmé, et est utilisé dans la thèse pour éliminer une hypothèse supplémentaire dans la démonstration principale. Je publierai cet argument-paradoxe en 1988. Un américain, Tim Maudlin, publiera une argumentation conceptuellement équivalente en 1989. La preuve de Maudlin est plus informative que la mienne. L'argument du graphe filmé est exposé au chapitre 4 de la thèse.

Après mes heures de service d'objecteur de conscience à l'ULB, je donnerai un cours portant sur le langage de programmation fonctionnelle LISP et sur la programmation logique en PROLOG ainsi qu'une introduction à l'Intelligence Artificielle[10], au calcul lambda, aux combinateurs et aux réseaux de neuronnes.

Ces cours auront un certain succès à l'exception des informaticiens de l'ULB qui ne voulaient entendre parler ni d'intelligence artificielle, ni de programmation logique ou fonctionnelle.

À la VUB par contre, la *Vrij Universiteit van Brussel"*, la version néerlandophone de l'Université Libre de Bruxelles, Luc Steels est de retour du MIT Américain[11], avec de l'argent, des projets en Intelligence Artificielle et des machines LISP. Je travaillerai alors à la VUB, je ferai quelques conférences, en flamand, sur les "capacités introspectives" des machines universelles et son application possible en Intelligence Artificielle, et je me verrai proposer une place d'assistant en informatique. Bien que je perfectionnerai mon néerlandais, que j'avais appris à l'école, à raison d'un cours intensif de quatre heures par jour pendant trois mois, je me verrai refuser la place d'assistant parce que j'avais encore un accent francophone trop prononcé ! Luc Steels n'est en rien responsable ; c'est le président de la faculté qui craignait que mon accent francophone ne pollue l'esprit des étudiants flamands. Je me heurtais ici à ce triste et fameux problème communautaire belge.

Je mentionne que j'avais travaillé aussi, avant le service civil, au service du professeur de mathématiques de la faculté de psychologie, Monsieur Ducamp, un des premiers, avec un ingénieur, Pierre van Nypelseer, à s'être interessé à l'intelligence artificielle à l'ULB. Cela me permettra d'avoir un compte sur l'ordinateur de l'université.

---

vol. 6, n2, aprile 1983.

[10]J'apprécie l'Intelligence Artificielle, aussi bien connexionniste (neuronale) que symbolique. Je crains cependant que l'expression "Intelligence Artificielle" soit un peu malheureuse. Si on admet d'entendre le mot "artificiel" par "introduit par l'homme", on comprend que la distinction entre naturel et artificiel est ... *artificielle*.

[11]Massachusetts Institute of Technology.

Après le service civil, on me proposera de travailler pour une société privée "Plant Genetic System" une société flamande de biotechnologie qui siège dans la ville de Gand, et qui investissait dans un service de recherche de l'ULB : l'Unité de Conformation des Macromolécules Biologique" (UCMB), dirigée par la biochimiste Soshana Wodak.

Comme on peut le constater, je suis toujours à l'ULB. Avant le service civil j'ai enseigné dans les écoles du secondaire mais j'avais un bureau au service de Cosmologie quantique du professeur Englert. J'ai aussi de bonnes relations avec le département de psychologie. Je serai par ailleurs régulièrement invité aux séminaires interdisciplinaires organisés par les psychanalystes pour faire des exposés de logique et de topologie élémentaire, et même plus tard pour exposer mes propres travaux. Ensuite j'effectue mon service civil d'objecteur de conscience à l'ULB, où mon bureau cotoie le bureau de X (!). Les rares fois où il viendra à son bureau, il ne m'adressera pratiquemment jamais la parole. Ensuite l'UCMB, puis plus tard l'IRIDIA. En fait c'était normal, outre que j'ai besoin du contact avec les chercheurs et les étudiants, j'avais développé un certain savoir-faire en logique et en intelligence artificielle, et ce domaine était très vivant. Bien que délaissées par la faculté des sciences—et par la faculté de philosophie et lettres—il y avait un réel besoin suscité par ces nouvelles techniques.

À l'UCMB je rencontrerai Michel Bardiaux, collègue et puis ami, ingénieur passionné par le langage ADA, qui me poussera à développer un interpréteur PROLOG—de programmation par la logique. J'écrirai un interpréteur PROLOG en LISP,en 15 jours, et pendant un an Michel et moi le traduirons et l'optimiserons en ADA afin de l'utiliser pour automatiser des raisonnements élémentaires sur la structure des protéïnes. Ce sera une expérience formidable et très enrichissante pour moi.

Inspiré par les travaux d'Ehud Shapiro[12] sur la correction automatique des programmes, je commencerai à développer un système, ANIMA, capable d'apprendre par une technique d'autocorrection. Le programme se corrige perpétuellement lui-même à différent niveaux d'autodescription. C'est ce travail qui va me lancer plus profondément dans l'informatique théorique et me fera revenir concrètement à la logique mathématique, notamment à l'analyse par la logique modale du théorème de Gödel et plus généralement de l'autoréférence. Je reprendrai mes pélerinages Carrolien à Oxford, et j'achèterai, en 1986, les livres de Georges Boolos "The Unprovability of consistency" (1979) et de C. Smoryński "Self-Reference and Modal Logic" (1985). Je

[12]Algorithmic Program Debugging, The MIT Press, 1983.

développe alors un profond intérêt pour la logique modale, et je recouvre pratiquement tout mon intérêt pour les phénomènes gödéliens d'incomplétude.

Le travail à l'UCMB prenait du temps et en ce qui concerne la "thèse", celle-ci était toujours considérée comme un hobby qui interférait, cependant, avec la profession.

L'histoire de la thèse peut être décrite comme un lent retour à la pureté cristalline de mes investigations enfantines, où la distinction entre la part communicable et incommunicable du secret de l'amibe est clarifiée par les conséquences des phénomènes d'incomplétude Gödéliens en informatique théorique.

Néanmoins la chimie quantique va jouer un rôle indirect mais capital. Pour expliquer ce rôle, je vais revenir brièvement aux années 1977 et 1978.

Dans l'histoire de la thèse, le passage de la chimie (quantique) est *logiquement* un détour, très utile néamoins pour comprendre et motiver le *résultat* du travail.

La physique en général et la mécanique quantique en particulier vont jouer continuellement un rôle indirect dans cette recherche. La physique ne pouvait pas être le point de départ car le physicien tient pour acquis ce dont je cherchais l'explication : l'univers ou l'apparence de l'univers, les lois physiques ou l'apparence des lois physiques. La mécanique quantique sera pour moi plus une cible visée qu'une base sur laquelle construire une théorie.

En 1977, le biologiste gödélien était mort ou sérieusement assommé. Vive le chimiste donc, vive le physicien peut-être. Ce n'est pas l'amibe qui *se* divise, c'est *l'univers* qui divise l'amibe, et l'amibe ne fait rien, elle n'existe plus. Morte et enterrée.

Mais qu'est-il donc cet univers ? Est-il vraiment fait de quelque chose et alors qu'est-ce ? La matière m'a toujours semblé plus élusive que la vie et la conscience.

Je relirai alors rapidemment le Linus Pauling, pour arriver illico au Cohen Tannoudji Diu Laloë, promu instantanément en nouveau "Watson", rapidemment accompagné de deux formidables "d'Espagnat" : "La Physique Contemporaine" et "Conceptual Foundation of Quantum Mechanics". Je relirai attentivement les adorables livres de Louis de Broglie, que j'avais entrepris pendant la dernière année de mes études sans doute pour fixer mon attention ailleurs.

Je demanderai à un ami physicien ce qu'il en était de cet électron qui avait l'air de pouvoir passer par deux trous à la fois.

Ah, si seulement l'amibe était encore là ! Peut-être aurais-je pensé, simplement, que l'électron, à la façon de l'amibe, se duplique et passe par les deux trous. Oui mais l'électron peut passer par les deux trous et refusionner ensuite. Ah ! Mais des fois les cellules fusionnent aussi, comme les spermatozoïdes et les ovules—qui ne sont pas des protozoaires de nos eaux douces, j'espère que vous êtes au courant !. L'analogie entre l'électron qui passe par deux trous et l'amibe qui se divise (ou se multiplie[13]) est naïve et carrément boiteuse, mais derrière elle se cache une idée clé qui mettra du temps à se dégager de mon esprit[14].

L'ami physicien me donne les références de l'article d'Einstein Podolski Rosen et celui-ci me procure une grande surprise.

Comme cette surprise est capitale pour motiver le développement de la thèse, revenons un moment sur la question "qu'est-ce que comprendre ?" dans l'esprit de certains physiciens.

On se souvient de la discussion que j'avais eue en 1972-73 avec mon ami Dominique : je n'étais pas satisfait avec l'idée des physiciens qu'un ensemble d'équations puissent servir d'explication. Prédire n'est pas expliquer, comme le dira si bien René Thom. Avec les équations, quand elles ne sont pas trop compliquées, on peut prédire les phénomènes. Mais en soi l'équation n'explique rien. Elle compresse, certes de façon très ingénieuse, la description du monde physique mais elle n'explique pas la nature des corps ni pourquoi il y a des corps, ni pourquoi ces corps obéissent à des lois, ni d'où viennent ces lois.

Avec l'article d'Einstein, Podolski, Rosen, je réalise que la théorie quantique est beaucoup plus étrange que je ne l'avais cru. Il met en quelque sorte explicitement le physicien en flagrant délit de parler sans avoir une idée précise de ce dont il parle. Il rend incontournable le problème de l'interprétation de la mécanique quantique en particulier et de la physique en général. Il annonce aussi le travail de J. S. Bell de 1964 : des étrangetés quantiques peuvent être testées expérimentalement. Des propositions "métaphysiques" entre dans le laboratoire pour y être testées !

La théorie quantique, dans sa formulation usuelle, décrit *deux* sortes d'évolution d'un système physique :

1. L'équation de Schrödinger. Elle décrit l'évolution du système physique

[13]Il est amusant de pouvoir user des deux formulations. Une raison profonde apparaîtra plus loin : la différence entre les discours de la première personne et de la troisième personne.

[14]En gros l'idée, très *Borgésienne*, que les histoires ou les calculs se multiplient et fusionnent. Cela sera précisé plus loin.

quand on ne l'observe pas. En gros l'équation décrit l'évolution tout à fait déterministe d'une onde, laquelle décrit les résultats possibles de l'observation.

2. Le principe de réduction de l'onde : lorsqu'on mesure une certaine grandeur, l'onde se réduit de façon non déterministe. La probabilité de telle ou telle autre réduction possible de l'onde est donnée par le carré de la grandeur de l'onde.

Par exemple, l'onde associée à une particule dont on vient de mesurer la position est donnée par une onde sphérique qui "diffuse la probabilité" de rencontrer la particule dans toutes les directions de l'espace. Si on répète ultérieurement la mesure de la position de la particule, on la trouvera pratiquement n'importe où.

Jusqu'à l'article d'EPR on expliquait ce phénomène en invoquant une perturbation due à l'appareil de la mesure. Cette idée est naturelle vu que les instruments du laboratoire sont en général beaucoup plus gros que la particule observée.

Ce qu'Einstein, Podolski et Rosen ont expliqué, c'est que si on prend au sérieux la mécanique quantique, la pertubation ne peut pas être mécanique ou physique au sens habituel du terme. En effet, disent-ils, la façon dont les ondes sont associées aux "objets" physiques entraîne que si deux particules interagissent alors elles ne sont plus décrites que par *une* seule onde. En effectuant une mesure sur une des deux particules, on réduit l'onde unique décrivant les deux particules et ce faisant on perturbe instantanément l'autre particule. Aucune définition raisonnable de la réalité ne peut admettre une telle forme d'inséparabilité, selon Einstein, et donc la mécanique quantique est fausse ou gravement incomplète.

Avec l'article d'Einstein Posolski Rosen (EPR) l'idée qu'une équation est une explication est battue en brèche : il faut encore que l'équation décrive une réalité, quoi qu'elle soit, intelligible, et l'article EPR illustre clairement qu'avec l'équation de Schrödinger, cela est loin d'être évident.

En 1964 Bell montrera que l'on peut tester expérimentalement l'existence de cette non séparabilité, ce qui conduira à une série de nombreuses expériences culminant avec celle d'Aspect à Paris en 1981. Elles confirmeront la mécanique quantique, contra Einstein & Al, mais dans un certain sens, elles confirmeront la prédiction de la mécanique quantique, qu"Einstein & Al ont épinglée, de l'existence de "perturbation à distance". On dit aujourd'hui que les états des deux particules sont enchevêtrés (entangled, en anglais).

On ajoute souvent que cette pertubation instantanée est aléatoire si bien qu'elle ne permet pas de transmettre de l'information instantanément. C'est exact. Malheureusement beaucoup en déduisaient que les états enchevétrés

ne pouvaient pas avoir d'applications. On sait aujourd'hui qu'il n'en est rien. Depuis les travaux précurseurs de Feynman, puis ceux de David Deutsch 1985, on sait qu'il est possible d'exploiter et de gérer les états enchevêtrés des particules, comme avec l'ordinateur quantique ou avec le phénomène de téléportation quantique. Mais ceci nous entraîne un peu loin.

Il n'y a pas d'unanimité chez les physiciens au sujet de la façon dont on peut interpréter la mécanique quantique. On peut en gros distinguer deux familles.

- Ceux qui pensent que lors de la mesure il y a une "réelle" réduction de l'onde. Il y a une onde et l'observation de celle-ci provoque une réelle réduction physique de cette onde. Par exemple si l'onde d'un électron passe à travers deux trous, l'observation d'un trou transforme l'onde de l'électron en une onde d'un électron qui passe par un trou.
- Ceux qui pensent qu'une telle réduction n'existe pas. La théorie quantique est réduite à l'équation de Schrödinger, auquel l'observateur lui-même obéit. L'onde de l'observateur s'enchevêtre avec l'onde de l'électron créant une onde décrivant deux observateurs observant chacun l'électron passant par un trou.

Les premiers ont beaucoup de difficultés à combiner l'évolution déterministe dictée par l'équation de Schrödinger avec la réduction aléatoire. La solution proposée par von Neumann et Wigner consiste à attribuer un rôle spécial à la conscience. Les objets physiques obéissent à l'équation de Schrödinger, la conscience réduit l'onde.

Les seconds sont obligés d'expliquer l'apparence d'une réduction. C'est ce que Everett a commencé à faire en 1957. Il montre que si on applique l'équation de Schrödinger, non pas au système physique isolé, mais au couple constitué du système observé *et* de l'observateur, considéré comme une machine à mémoire, alors l'équation de Schrödinger seule arrive à prédire l'observation d'une réduction de l'onde *dans le discours des observateurs-machines*, dans le compte-rendu de leurs (suites) d'expériences. L'avantage est de justifier l'apparence de "pertubation à distance" et l'apparence d'indéterminisme dans un contexte globalement local et déterministe. Ce genre d'approche sera considérablement étendu dans mon travail, et même généralisé à la vérité arithmétique toute entière.

On peut consulter l'annexe sur la mécanique quantique dans la thèse pour un peu plus d'informations.

L'équation de Schrödinger appliquée au couple observateur/chose-observée prédit que l'état quantique de l'observateur s'enchevêtre avec l'état quantique

de la chose observée. Si l'onde de l'électron décrit la position de l'électron comme passant par les deux trous, l'onde de l'observateur va s'enchevetrer avec l'onde de l'électron avec comme résultat que l'onde globale décrit un état avec l'observateur observant l'électron dans un trou *et* l'observateur voyant l'électron dans l'autre trou. L'observateur s'est multiplié lui-même en observant l'électron. Il n'y a pas eu réduction de l'onde, même si du point de vue de chaque observateur, c'est tout comme.

Que les observateurs se multiplient sur l'arbre des éventualités possibles aurait du plaire au biologiste gödélien, amateur de Lewis Carroll et de Borges[15]. Mais en 1977-78 l'amibe m'était sortie de la tête et ce n'est que lentement que j'approfondirai la relation profonde qui existe entre le "mécanisme de Turing" (le computationnalisme) et la formulation d'Everett de la mécanique quantique. À cette époque, contrairement aux impulsions de mon enfance, je voulais croire à une matière toute mystérieuse. Ce n'est qu'avec le retour de mon intérêt pour Gödel et la découverte de la thèse de Church (voir le chapitre suivant) que je me rappelerai que les machines et les nombres portent suffisamment de mystères en eux et qu'il n'y a pas de raison d'en rajouter.

Note. J'ai finalement retrouvé mes carnets de 1981 à 1986. Cela m'a permis d'un peu mieux comprendre l'évolution des idées. En gros, comme je l'ai déjà dit, le biologiste Gödélien est assommé (disons). Il est clair que c'était le théorème de Gödel qui, en 1971, m'a fait entrevoir la possibilité de communiquer (prouver) le renversement et le fait que la physique et la chimie pouvaient de façon ultime reposer sur une base immatérialiste faite des rêves possibles des amibes abstraites. En apprenant que le théorème de Gödel n'était plus à la mode chez les logiciens (l'"erreur" de X), j'ai du inconsciemment abandonner cette conception des choses. Vers la fin de mes études et après, le chimiste en moi, matérialiste, est revenu. Celui-ci a hésité essentiellement entre deux interprétations de la mécanique quantique. D'une part celle de von Neumann ou Wigner, quasi dualiste où la conscience agit sur la matière en réduisant l'onde quantique. L'onde quantique décrit des histoires multiples et incompatibles, l'observation consciente 'choisi' une histoire parmi les histoires décrites par l'onde (cf l'électron qui passe par les deux trous). Et, d'autre part l'interprétation d'Everett, où il n'y a pas de réduction d'onde et où toutes les histoires sont "physiquement réalisées". Les observateurs, chez Everett, peuvent être considérés comme des machines à mémoires. Le sentiment qu'ils ont de l'unicité de leur propre histoire est due au fait qu'obéissant eux-mêmes à l'équation de Schrödinger, ils sont eux-mêmes multipliés. Je concevais ces deux interprétations comme étant matérialistes (au sens faible où je l'entends) : avec von Neumann-Wigner il s'agit de dualisme ou de double matérialisme, une conscience *substantielle* et n'obéissant pas à l'equation de Schrödinger, agit sur une matière substantielle qui y obéit. Cette interprétation soulève énormément de difficultés conceptuelles et techniques et je l'abandonnerai pour celle d'Everett. J'accepterai l'hy-

[15]Voir la nouvelle "Le jardin aux sentiers qui bifurquent", dans son recueil "*Fictions*".

pothèse du mécanisme, mais, toujours influencé par le chimiste en moi, je concevrai celle-ci encore dans le cadre du matérialisme. Il existe alors un super-univers au sein duquel notre univers matériel se multiplie. C'est d'ailleurs l'interprétation usuelle du formalisme de la mécanique quantique selon Everett. En 1984, je trouve le paradoxe du graphe filmé. Pour le paradoxe RE (paradoxe du déployeur universel) je n'ai pas de date. Je voyais ces paradoxes comme des sérieuses difficultés du mécanisme. Et, restant matérialiste, je les voyais presque comme des arguments en faveur du dualisme à la Wigner. Ce n'est que dans un carnet de 1986 que je réaliserai que le paradoxe RE, où toutes les histoires computationnelles sont générées, généralise l'explosion des histoires de l'univers comme cela semble être le cas avec l'interprétation d'Everett. Les solutions de l'équation de Schrödinger étant calculables, il s'agit d'une généralisation mathématiquement bien définie : les histoires multiples d'Everett sont des cas particuliers des histoires computationnelles multiples et immatérielles. Je réalise alors que les bizarreries quantiques pourraient très bien être une confirmation du mécanisme, car les bizarreries mécanistes mises en évidence avec le graphe filmé et avec le paradoxe RE ne sont pas plus étranges que les bizarreries quantiques, et ont de surcroît un air similaire. J'interpréterai alors le paradoxe du graphe filmé et le paradoxe RE, non plus comme une réfutation possible du mécanisme mais comme un argument pour le renversement, revenant ainsi à mon intuition de 1963. Entretemps mon intérêt pour le mécanisme et finalement pour Gödel étant revenu (sans nausée grâce au cannabis) je concluerai que le discours de la machine universelle autoréférentiellement correcte doit converger vers les discours physique et chimique, revenant ainsi à mon intuition de 1971. Je modéliserai la communication scientifique entre machines par la prouvabilité formelle, conformément à mon "intention" de 1971. Je mettrai encore du temps à penser à utiliser les théories de la connaissance du théétète de Platon pour capturer formellement les notions de première et de troisième personne, ainsi que la modélisation du déployeur universel par les formules Σ_1 (voir le chapitre sur le renversement, et le chapitre sur la machine et son ange gardien).

Je réalise surtout en feuilletant ces carnets que j'ai abandonné le *renversement* parce que je préférais croire m'être trompé plutôt que de croire que l'université m'avait trompé. Plus tard je me dirai aussi que Gödel, surtout à Princeton aux cotés d'Einstein, avait sûrement pensé à ce renversement, et que, s'il ne l'avait pas exploité, c'est qu'il avait compris que cela ne fonctionnait pas. Encore maintenant je trouve un peu étonnant que Gödel et Einstein ensemble n'aient pas découvert l'interprétation d'Everett de la mécanique quantique, et qu'ils n'aient pas découvert l'interprétation *à la Everett* de l'arithmétique, c'est-à-dire exactement ce que j'ai rendu logiquement obligatoire dans le présent travail, dès qu'on postule le computationnalisme. Il est vrai que Gödel n'a pas trop apprécié la thèse de Church (voir chapitre suivant), ni le mécanisme.

Je réalise qu'il y a beaucoup d'autres faits que je ne mentionne pas. Une preuve que le biologiste est revenu plus rapidemment que le biologiste gödelien est donnée par le fait que j'éleverai de 1982 à 1985 des *planaires*. Il s'agit d'un petit ver d'eau douce qui est un véritable champion de la régénération cellulaire. Il va m'inspirer pour mon approche théorique et gödélienne de la régénération et de la différentiation cellulaire. Je renvoie à mon article de 1992, "Amoeba, Planaria and Dreaming Machine". La planaire jouera un rôle important pour ma réflexion sur la biologie théorique. Un beau livre sur les invertébrés est le Ralph Buchsbaum : "Animals without Backbones : 1", Pelican books, 1938. Suivis de nombreuses rééditions corrigées et élargies, il contient un chapitre détaillé sur la régénération cellulaire.

Chapitre 6

La machine universelle retourne sur terre

Monsieur Bamberger, préfet de l'Athénée Maïmonide, n'en croyait pas ses oreilles. Voilà une demi-heure que votre serviteur gesticulait dans toutes les directions en tentant de le convaincre de réunir des fonds pour acheter des ordinateurs pour l'école, à l'usage des élèves. Jusque-là j'invitais les plus intéressés parmi ceux-ci à venir chez moi s'entraîner à la programmation récursive avec une *tortue* graphique ou avec des listes dans un langage LOGO que j'avais implanté en BASIC sur mon TRS 80.

L'école n'était pas riche, le bâtiment était à la limite de l'insalubre et en éternelles réfections qui nécessitaient des frais continus. Je ne me faisais pas trop d'illusions pour ces fonds.

'Et que feront les élèves avec des ordinateurs ?' me demande le préfet.

'Mais, monsieur le Préfet, pensez-y ! C'est un miroir dynamique qui excitera davantage les cellules grises de nos élèves, c'est un accélérateur universel, un tunnel sur d'autres mondes, un trou noir épistémologique. C'est la machine philosophale que l'on peut questionner, c'est la machine que l'on peut mettre en mouvement avec le verbe, c'est le Golem, Monsieur le préfet, et peut-être bien plus !' Je me laissais aller parce que je me sentais en confiance.

'Monsieur Marchal, j'apprécie votre enthousiasme bien que je pense que vos propos sont un peu exagérés. Mais vous savez que l'école n'est pas riche, le bâtiment est à la limite de l'insalubre et en éternelles réfections qui nécessitent des frais continus. Nous n'arrêtons pas de faire des fancy-fair[1]. Aussi, si vous voulez bien, nous en reparlerons une autre fois, je dois m'en aller.'

Et par une de ces coincidences qui ne doivent rien au hasard, Monsieur

[1]Terme anglo-belge désignant une sorte de boum souvent lucrative.

Bamberger s'en alla en Israël visiter des écoles "pilotes" bourrées d'ordinateurs, justement.

À son retour il me convoqua dans son bureau et me donna carte blanche pour introduire des ordinateurs à l'école. Je donnerai un cours d'informatique aux instituteurs et aux élèves. J'organiserai un club informatique, et de nombreux sympatisants, à l'école, organiseront les fancy-fair et la quête des parents généreux pour acheter 5 magnifiques APPLE 2. A Maïmonide, Monsieur Bamberger me considéra comme un prophète.

Prophète ? Moi ? Je ne sais pas si je suis un prophète, mais *si* j'étais le prophète de la machine universelle, *alors* je serais un prophète assez maladroit. Déjà que je n'ai pas reconnu la bête, la prenant pour de stupides frigos rigides servant de mémoires à l'usage exclusif des banquiers.

Surtout j'arrive en retard.

J'aurais dû venir au siècle précédent annoncer la machine de Babbage. Celui-ci concevra et commencera à construire, en Angleterre, une machine universelle, toute de roues dentées et de clapets métalliques. Son fils terminera la construction de la machine. Après quoi la machine ira au musée de Londres où on peut toujours la voir fonctionner. Le petit livre visionnaire de Jacques Lafitte (1911) "Réflexions sur la science des machines[2]" stipule que Babbage aurait plus souffert de l'incompréhension de ses contemporains pour son invention d'un système de notations fonctionnelles que pour sa machine. Cette notation lui servait à décrire son fonctionnement, et il a dû réaliser leur équivalence computationnelle, ce que je prends comme indice qu'il a pressenti l'universalité de sa machine et de son système de notation. Et, donc, je vais l'expliquer, qu'il a pressenti la thèse de Church.

La thèse de Church affirme en effet que tous les ordinateurs possibles, qu'ils soient matériels ou virtuels, comme les langages de programmation, sont équivalents. Ils sont équivalent dans le sens, qu'abstraction faite du temps d'éxécution, ils sont à même de s'émuler (de simuler parfaitement) les uns les autres. Ils définissent tous la même collection de fonctions calculables. Je reviens là-dessus plus loin.

Si j'avais été prophète de la machine universelle, j'aurais dû annoncer la machine universelle de Turing. Elle apparaît, ce siècle-ci, à partir d'une réflexion sur les fondements des mathématiques suite à la crise cantorienne des mathématiques au début du siècle[3]. Elle est la base de l'informatique

[2]Éditions Vrin, 1972.

[3]J'approfondis ce point dans "Conscience et Mécanisme", RT Iridia, 1995.

théorique du siècle, et Turing énoncera la "thèse de Church" dans son papier de 1936 où il définit et démontre, le premier, l'existence de la machine universelle : cette machine capable d'imiter toutes les autres machines. Bien que véritable héros de la seconde guerre mondiale, et malgré ses variés et importants travaux, de la chimie théorique à la logique mathématique et à l'informatique théorique et pratique, y compris l'intelligence artificielle, les réseaux neuronaux et les fondements de la mécanique quantique, il ne sera pas trop reconnu de son vivant. Il sera emprisonné pour homosexualité. Il mettra fin à ses jours.

Ou alors, excusez-moi de retourner une quinzaine d'années en arrière, j'aurais dû annoncer les systèmes normaux de Post. En 1921, dans une étonnante anticipation, Emil Post, aux États-Unis, invente ou découvre (comme vous voulez) un système symbolique, aujourd'hui on dit un système formel, universel. À partir d'une réflexion logique sur les méthodes de manipulations finies, il énonce la "thèse de Church" 20 ans avant tous les autres (Turing, Kleene, Markov) sous la forme d'une loi de l'esprit. Il dérive (non constructivement) à partir de cette 'loi de l'esprit' une forme générale du théorème d'incomplétude de Gödel (10 ans avant Gödel[4]). Il découvre l'argument, fondé sur le théorème "de Gödel", montrant que l'homme est supérieur à la machine (38 ans avant Lucas, 68 ans avant Penrose). Il découvre *l'erreur* dans cet argument (59 ans avant Webb, avant Benacerraf, et de plus en plus d'autres ...), etc. Post est surtout connu des logiciens pour avoir énoncé, dans son brillantissime article de 1944[5], un fameux problème qui sera résolu indépendemment aux USA et en URSS et qui sera la base de la théorie de la récursion. C'est vraiment une théorie de la diagonalisation, en fait. C'est aussi la source d'inspiration d'une très grosse part de l'informatique théorique. Dans ses notes de 1921 il en arrive à considérer le monisme immatérialiste[6], mais en fait, à cet endroit, dans une note en bas de page, il affirme avoir changé d'avis et être revenu, en 1924, au dualisme, influencé semble-t-il par Turing.

[4]L'incomplétude gödélienne est une conséquence assez directe de la thèse de Church. Voir l'annexe sur la thèse de Church pour une explication détaillée de ce fait.

[5]Recursively Enumerable Sets of Positive Integers and their Decision Problems. American Mathematical Society, 1944, Volume 50, pp. 284-316. L'article, comme ceux de Gödel sur l'incomplétude, de Kleene, Church, Rosser sont repris dans ce Super-Watson qu'est la sélection d'articles par Martin Davis : The Undecidable. Référence dans la thèse. L'anticipation des années 1920 de Post est dans le Davis, et nulle part ailleurs, à ma connaissance.

[6]La présente thèse montre que le computationnalisme entraîne le monisme immatérialiste, c'est-à-dire la doctrine idéaliste selon laquelle la matière émerge de l'esprit et non l'inverse. L'esprit, ici, est défini exclusivement par la vérité mathématique ou même seulement arithmétique, y compris les vérités autoréférentielles relatives quelles soient prouvables ou non.

Ou alors j'aurais dû annoncer les algorithmes de Markov ? On peut démontrer qu'une fonction est calculable par un algorithme de Markov si et seulement si elle est calculable par une machine de Turing. Et Markov, indépendemment de Turing et de Post, énonce, en URSS, la "thèse de Church".

N'aurais-je pas dû, toujours à supposer que je sois le prophète de la machine universelle, annoncer les combinateurs de Curry, les fonctions lambda de Church et finalement l'ordinateur concret de von Neumann, et tous les langages de programmation qui sont apparus depuis et qui définissent tous des machines virtuelles mais tout autant universelles ? Après tout le prophète de la machine universelle annonce celle-ci, et pas nécessairement seulement ceux qui ont pris conscience de la portée de l'universalité en énonçant la thèse de Church, ou une thèse équivalente.

OK, mais alors, n'aurais-je pas dû annoncer l'invention du téléphone par l'amibe, pardon, je veux dire l'apparition du système nerveux biologique chez les animaux. Après tout le cerveau, qui nous permet de rêver à ces machines universelles et même aujourd'hui d'en construire, est assurément *au moins* universel[7]. Il n'est en effet pas difficile de vous convaincre que vous-même, lecteur, êtes capable, si on vous donne le temps et l'espace, d'émuler un ordinateur ; peut-être devrais-je *vous* annoncer ?

Et pourquoi n'aurais-je pas dû annoncer l'apparition des circuits moléculaires de régulation génétique qui, manifestement, ont été capables de soutenir l'émergence du cerveau et de votre personne. Il semble que si j'étais prophète de la machine universelle, je devrais annoncer son éternel retour.

En réalité c'est bien la thèse de Church qui fait de la machine universelle *de Turing* une machine universelle *tout court*. Et dans notre histoire humaine, la thèse de Church témoigne non seulement de la (ré)apparition de la machine universelle, mais surtout consacre notre prise de conscience de son universalité ainsi que du caractère épistémologiquement absolu de la notion de calculabilité par des procédures finiment descriptibles.

Church lui-même, curieusement, ne proposera pas *la thèse de Church*. Il proposera simplement de définir la notion de fonction calculable par la notion formelle de fonction lambda-calculable. On n'a pas besoin de comprendre ce que cela signifie précisément pour comprendre la suite de l'histoire. Il suffit de comprendre que Church, comme Turing et les autres, propose une définition

[7]Ceci est vrai indépendemment de l'hypothèse du computationnalisme. Le computationnalisme est l'hypothèse que nous ne sommes pas plus qu'une machine universelle, dans le sens où on suppose qu'une machine universelle est suffisante pour nous émuler.

formelle de calculabilité. En effet, dès que Church propose sa définition, Kleene, n'y croit pas. C'est un peu absurde de ne pas croire à une *définition* évidemment. Disons que Kleene ne croit pas que la définition de Church est adéquate. Il ne croit pas qu'on puisse même donner une définition à la fois formelle et absolue de la notion de calculabilité. Kleene connaissait bien le résultat de Gödel qui montre que la notion de prouvabilité formelle est relative. En effet Gödel avait démontré, comme je l'ai esquissé plus haut, que l'ensemble des propositions prouvables dans une théorie formelle n'était pas fermé pour la diagonalisation. C'est-à-dire qu'avec la diagonale on peut mettre en évidence des propositions vraies, exprimables dans la théorie formelle, mais néanmoins indémontrables dans cette théorie formelle. Stephen Cole Kleene était persuadé que toute notion de calculabilité formelle, en particulier le formalisme lambda de Church, doit lui aussi subir le joug de la diagonalisation.

Kleene s'était cru capable de critiquer la définition de Church en produisant, par diagonalisation, une fonction intuitivement calculable qui ne serait pas calculable dans le système de Church, réfutant ainsi la prétention à l'universalité du formalisme de Church. Kleene raisonne ainsi. Comme le système de Church est formel, cela entraîne que chaque définition de fonction calculable particulière dans ce système est représentable par une expression formelle grammaticalement bien définie. On peut donc décider "mécaniquement" si une expression formelle du système de Church représente une fonction calculable. Mais alors on peut ranger méthodiquement toutes les fonctions calculables définissables dans le système de Church. Il suffit de les ranger par leur longueur (définie par le nombre de signes dans l'expression formelle), et de ranger celles qui ont la même longueur par ordre alphabétique. On obtient alors la liste de toutes les fonctions lambda calculables, qui, si on juge la définition de Church adéquate devrait donner *toutes* les fonctions calculables :

$$f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7, ...$$

Considérons alors la fonction g définie au moyen d'une *première diagonalisation* :

$$g(n) = f_n(n) + 1$$

La fonction g est clairement *intuitivement, mécaniquement*, calculable. Pour calculer g appliqué à n, il suffit d'aller chercher dans la liste—qui est elle-même générable mécaniquement—la nième fonction f_n, de l'appliquer à n et d'ajouter 1.

Mais la fonction g ne peut *pas* être définie par une expression lambda du système formel de Church ! En effet, si c'était le cas, elle appartiendrait à la liste. Et donc il existerait un nombre k tel que $g = f_k$, et une deuxième diagonalisation pourrait se mettre en branle. En effet si on applique g à son propre numéro dans la liste, k, on a évidemment que $g(k) = f_k(k)$. Mais par définition de g, on a aussi :

$$g(k) = f_k(k) + 1$$

Ce qui reviendrait à dire que :

$$g(k) = g(k) + 1$$

Et en soustrayant $g(k)$ à gauche et à droite du signe =, on obtient :

$$0 = 1$$

N'est-ce pas là une formidable démonstration par l'absurde de l'incomplétude du système formel de Church, et même de tout système formel prétendant définir toutes les fonctions calculables ?

Eh bien non ! En l'occurence on peut montrer que la fonction g est parfaitement définissable dans le système de Church. Que se passe-t-il alors lorsqu'on calcule $g(k)$, c'est-à-dire g appliqué à son propre numéro dans la liste ? On obtient que $g(k)$ n'est pas définie. En terme d'éxécution par une machine, on obtient une éxécution infinie, et il n'y a rien de bizarre à ce que l'infini soit égal à l'infini +1.

La démonstration de Kleene montre seulement qu'il n'est pas possible de générer mécaniquement la liste de toutes les fonctions calculables *qui soient en même temps* parfaitement définies sur tous leurs arguments. Dès qu'on admet dans la liste des fonctions qui ne seraient pas définies pour certaines valeurs, alors rien ne s'oppose à penser que la liste contient *toutes* les fonctions calculables, y compris toutes celles qui sont partout bien définies, parce que les diagonalisations ne conduisent plus à des contradictions. Kleene lancera alors le vocable "thèse de Church", en constatant que la classe des fonctions calculables, qui ne sont pas nécessairement partout définies, est fermée, comme on dit, pour l'opération transcendantale et cantorienne de diagonalisation. C'est presque un jeu d'enfant, si on a bien intégré le raisonnement de Kleene, d'obtenir le résultat de limitation de Gödel et bien d'autres, à partir de la thèse de Church[8]. Par exemple, il est évident qu'on ne pourra

[8]Dans ce sens le théorème de Gödel *confirme* la thèse de Church. Judson Webb, à ce sujet dit que le théorème de Gödel est *l'ange gardien* de la thèse de Church et du mécanisme.

jamais construire une machine capable de décider, à partir d'une description formelle d'une fonction, si celle-ci est partout définie. En effet, si on disposait d'une telle machine, on pourrait extraire mécaniquement de la liste plus haut une sous-liste des fonctions calculables *partout définies*, et avec la thèse de Church, on les obtiendrait toutes. Et cette fois-ci, avec cette sous-liste, le raisonnement de Kleene conduirait vraiment à montrer que $0 = 1$. Le prix de la généralité promise par la thèse de Church est donné par l'ensemble des résultats de limitations concernant les machines universelles. Leur universalité les rend imprédictibles, essentiellement incontrôlables. À vrai dire ces machines, plus on les étudie, plus on réalise qu'il s'agit de véritables inconnues. La thèse de Church protège les machines de toutes théories réductionnistes (complètes) que l'on pourrait inventer à leur sujet. Avec, *en plus*, l'hypothèse du computationnalisme, la thèse de Church *nous* protège des psychologies normatives ; elle épingle l'inconnu en notre propre sein. Ceci sera précisé au chapitre 8.

Et Gödel ? Gödel ne proposera pas la thèse de Church. En fait il n'y croira pas pendant un temps. Selon ses dires, ce n'est qu'après la lecture attentive de l'article de Turing qu'il commencera à admettre la thèse de Church[9]. Gödel prendra alors conscience de l'aspect quasi miraculeux de cette thèse. En 1946, à Princeton[10], il estimera que la fermeture de la classe des fonctions calculables pour l'opération de diagonalisation est une sorte de miracle :

> Tarski[11] has stressed in his lecture (and I think justly) the great importance of the concept of general recursiveness (or Turing's computability). It seems to me that this importance is largely due to the fact that with this concept one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e., one not depending on the formalism choosen.

[9] Il ne croira cependant jamais vraiment au computationnalisme. Voir le rapport technique IRIDIA 1995 pour plus de renseignements à ce sujet.

[10] Voir la référence à la fin de la thèse.

[11] Tarski a insisté dans sa conférence (et je crois qu'il a raison) sur la grande importance du concept de récursivité générale (ou de la calculabilité de Turing). Il me semble que cette importance est largement due au fait qu'avec ce concept on a pour la première fois réussi à donner une définition absolue d'une notion épistémologique intéressante, c'est-à-dire, une notion ne dépendant pas du formalisme choisi. Dans tous les autres cas traités avant, tels que la prouvabilité et la définissabilité, on a été capable de les définir seulement relativement à un langage donné, et pour chaque langage individuel il est clair que la notion obtenue n'est pas celle que l'on cherche. Pour le concept de calculabilité pourtant, quoiqu'il ne s'agisse que d'une sorte de prouvabilité ou de décidabilité, la situation est différente. Par une sorte de miracle il n'est pas nécessaire de distinguer des ordres, et la procédure de la diagonale ne nous conduit pas hors de la notion définie.

> In all other cases treated previously, such as demonstrability or definibility, one has been able to define them only relative to a given language, and for each individual language it is clear that the one obtained is not the one looked for. For the concept of computability however, although it is merely a special kind of demonstrability or decidability the situation is different. By a kind of miracle it is not necessary to distinguish orders[12], and the diagonal procedure does not lead outside the defined notion.

Gödel espéra en vain un "miracle" semblable pour la notion de prouvabilité. Mais avec son propre théorème d'incomplétude allié à la thèse de Church, on peut craindre que cela soit difficile. La notion de prouvabilité formelle est essentiellement relative à la différence de la notion de calculabilité. On verra cependant, quand on va interviewer la machine et son ange gardien (dans deux chapitres), comment on va arriver à formaliser indirectement une notion de preuve intrinsèquement non formalisable (!) et quasi-absolue, *du point de vue* de la machine. Mais cela est pour plus tard.

C'est la thèse de Church qui permet à de nombreux résultats d'informatique théorique d'être *machine-indépendants*. les résultats ne dépendent pas du choix du système formel utilisé. De la même façon qu'en géométrie les théorèmes importants sont ceux qui ne dépendent pas du choix d'un système de coordonnées, en informatique théorique, de même, le choix d'une machine universelle définit une sorte de base dans laquelle on peut identifier les machines avec des nombres, et cela de façon telle que les résultats ne vont pas dépendre du choix de la base.

Dans mon travail la thèse de Church garantit la généralité du déployeur universel (DU). Celui-ci est un programme, qui, non seulement est *capable* d'émuler toutes les machines digitales (numériques), mais qui les émulent effectivement toutes. De façon imagée, on peut voir le DU comme une machine universelle *écrasée*, dont jailliraient toutes les éxécutions possibles. Ce n'est pas difficile de transformer une machine universelle en déployeur universel.

La seule petite difficulté technique provient du fait de l'existence d'éxécutions infinies de certaines machines digitales (une conséquence, donc, de la fermeture pour la diagonalisation de la collection des fonctions calculables, comme le raisonnement de Kleene l'illustre précisément).

[12]A la différence de la prouvabilité formelle qui étant nécessairement relative—conséquence du théorème d'incomplétude de Gödel—peuvent s'étendre en escaladant ce que les logiciens qualifient souvent d'ordres ou de types.

Pour construire un déployeur universel, il suffit en effet de construire un générateur de tous les programmes acceptables par une machine universelle donnée et de zigzaguer (dovetail, en anglais) sur tous les morceaux d'éxécutions finies de cette machine. Il s'agit d'une technique bien connue en informatique théorique et découle du fait que le produit cartésien d'ensemble mécaniquement générable est mécaniquement générable[13]. Le DU généralise grandement la bibliothèque de Babel qui contient tous les livres, car il génére non seulement tous les livres, y compris ceux qui sont infinis[14], mais aussi, avec l'hypothèse du computationnalisme, toutes les lectures possibles de ces livres, et tous les rêves que ces lectures entraînent.

[13]Le terme consacré, en admettant la thèse de Church, pour "mécaniquement générable" est "récursivement énumérable", d'où l'appellation "paradoxe RE" pour le "paradoxe du dovetelleur universel", devenu finalement "argument du déployeur universel". "dovetelleur" vient du terme consacré en informatique théorique pour ce *zigzaguage* : dovetailing, qui lui-même est un terme utilisé par les toituriers pour décrire une façon de placer des tuiles, en queue d'aronde, sur un toit.

[14]Le DU génère donc tous les nombres réels. Il n'y a pas de contradiction avec le théorème de Cantor qui dit que l'ensemble des nombres réels n'est pas dénombrable, ce qui voudrait dire que l'on peut faire une liste de tous les nombres réels. Le DU ne génère pas de telle liste. Il génère chacun des réels petit-à-petit, sans jamais proposer une telle liste. L'algorithme suivant génère aussi chacun des réels à la façon du DU : générer $0,0$ et $0,1$ puis $0,00$ et $0,01$ et $0,10$ et $0,11$ puis les 8 suivants, puis les 16 suivants, etc. De la même façon le Déployeur universel génère tous les nombres aléatoires ou non-algorithmiquement compressibles, comme le fait le simple comptage usuel 0, 1, 2, 3 etc.

Chapitre 7

Le renversement ("1963" revisité)

Je réalise que ce serait long d'expliquer le cheminement de la thèse avec tous ses détours, et qui, comme j'ai dit plus haut, n'est en fait qu'un lent retour à la pureté cristaline de l'exposé de 1963 sur l'amibe.

Un *vrai* progrès est la découverte de la thèse de Church et de la machine universelle (voir chapitre précédent). Un *faux* progrès est sans doute le suivant. Le "résultat" de 1963 est "Si une amibe vit deux jours alors elle vit toujours". Je suis parti alors dans une quête obsédée d'une preuve que l'amibe vit deux jours (c-à-d survit à sa duplication). Aujourd'hui j'ai complètement (ré)intégré le fait que la réponse à cette question est *vraiment* incommunicable *à la troisième* personne. Ou si vous préférez est scientifiquement incommunicable. Mon père et le Ames & Wyler avaient raison de se taire, l'amibe, comme toute machine autoréférentiellement correcte, demeure silencieuse sur cette question. Mais l'amibe peut parier.

Le computationnalisme est une hypothèse : espérance ou crainte selon les goûts. La beauté est que le computationnalisme justifie son caractère incommunicable. Cela est justifié intuitivement avec l'expérience par la pensée de l'autoduplication de soi, et rendu formellement clair avec les interviews de la machine universelle et de son ange gardien (voir chapitre suivant).

La question n'est donc pas de savoir si le *mécanisme* digital ou numérique est vrai ou faux. La question est de savoir si vous acceptez une greffe de cerveau artificiel digital. La question a un caractère relativement urgent, car on a commencé à substituer des parties du cerveau par des artefacts électroniques. En particulier un aveugle a retrouvé une "sorte de vision" grâce à une telle substitution. Des progrès spectaculaires ont aussi été réalisés sur des animaux. Et la question n'est pas *vraiment* de savoir si vous allez survivre avec ce cerveau artificiel, mais seulement de comprendre que si,

effectivement, vous survivez, alors le renversement psychologie/physique suit nécessairement.

Il est naturel que les conséquences logiques de propositions incommunicables soient elles-mêmes incommunicables. Je montrerai dans le chapitre suivant comment le retour à Gödel, et à mon intuition de 1971[1], permettra d'isoler mathématiquements les parts communicables et incommunicables du discours de la machine universelle. Par communicable j'entends scientifiquement communicable, ou plus simplement comme je vais l'illustrer en détail, communicable *à une troisième personne* ou *prouvable*, par opposition à une autre forme de "communication" que je vais appeler *communication à la première personne*. La distinction entre première et troisième personne, qui va être expliquée ici, est un autre progrès, conceptuel et pédagogique, pour expliquer plus aisément le renversement[2].

Je vais m'inspirer de la version la plus récente de l'argument du déployeur universel (chapitre 3 de la thèse), celle que j'ai exposée la semaine dernière, en avril 2000, à Dubrovnik, au 26ème Congrès International de Philosophie des Sciences.

Remarque. La définition du computationnalisme présuppose un minimum de psychologie *folklorique*, ou de *grand-mère*, , comme on l'appelle parfois. C'est la psychologie de la vie quotidienne. En particulier il est nécessaire de savoir donner un sens au mot "survivre" dans un minimum de situations données.

Lorsque j'expose oralement le raisonnement, il m'arrive, pour illustrer le minimum de psychologie de grand-mère nécessaire, de commencer par proposer l'expérience concrète suivante. Je demande à l'assemblée si elle m'autorise à lâcher mon crayon sur le pupitre. En général, l'assemblée, un peu surprise, finit par acquiesser, et je laisse tomber le crayon, et il tombe alors, chaque fois. Et personne n'y trouve rien à redire. Je demande alors explicitement à l'assemblée si elle estime avoir *survécu* à cette expérience, et si elle survivrait alors que je réitèrerais l'expérience du lâcher du crayon. C'est juste pour illustrer que nous sommes à même de donner un sens commun au mot "survivre" dans le sens où nous admettons survivre aux 1001 événements quotidiens. Ensuite, pas à pas, je leur demande s'ils estiment possible de survivre à une greffe de cœur artificiel, ... jusqu'au cerveau artificiel. Par *définition*,

[1]Intuition selon laquelle la preuve d'incomplétude de Gödel illustre comment séparer du prouvable de l'improuvable pour des classes de discours formels.

[2]Au chapitre suivant, la communication à la troisième personne va être modélisé (ou même "capturé") par la prouvabilité formelle (arithmétisable, gödélienne). Les nuances du type "première personne", "troisième" personne" seront capturées par des variantes modales de la prouvabilité gödélienne, inspirée du théétète de Platon.

le computationnaliste est celui qui répond oui à toutes ces questions.

En ce qui concerne le passage du cœur au cerveau, j'entend souvent une objection : 'j'imagine que je peux survivre avec le cœur d'une autre personne, mais si on greffe dans mon crâne le cerveau d'une autre personne, c'est plutôt cette autre personne qui survit avec mon corps. Le cerveau et le cœur ont un rôle dissymétrique à cet égard'. Cette objection illustre deux choses. D'abord que la personne qui fait cette objection a une bonne intuition du mot survivre. La remarque est correcte et découle effectivement de l'hypothèse du neurophysiologiste selon laquelle le cerveau est l'organe de la mémoire et de la conscience. Cela illustre donc aussi que la greffe de cerveau artificiel doit impérativement être faite au bon niveau.

Si vous remplacez le chapitre 4 de "A la recherche du temps perdu" de Marcel Proust par le chapitre 4 d'"Alice au pays des merveilles", le contenu des livres est pertubé. Mais si vous faites le remplacement lettre par lettre en respectant les relations de proximité entre lettres, le contenu des livres sera invariant pour la substitution. Il en est de même avec le computationnalisme lors de la greffe de cerveau artificiel, la substitution doit être faite au bon niveau. Croire au computationnalisme, c'est croire que ce niveau existe. Nous verrons que ce niveau ne peut pas être déterminé avec certitude, amis il peut être parié correctement, comme on le supposera dans les expériences par la pensée.

Bien qu'il soit nécessaire d'admettre ce minimum de psychologie folklorique, celui-ci devra être éliminé pour permettre une extraction purement mathématique de la physique à partir de la psychologie des machines, ce qui est rendu possible par le chemin Gödélien, celui de 1971, comme je l'explique brièvement dans le chapitre suivant. L'idée consistera à substituer le discours de grand-mère par le discours gödélien de la machine autoréférentiellement correcte.

L'hypothèse précise du computationnalisme est la donnée des trois sous-hypothèses suivantes :

1. *Le pari mécaniste*. Il existe un niveau de description de moi-même tel que je survis à une substitution fonctionnelle, et digitalement descriptible, des parties qui me constituent à ce niveau. J'appelle un tel niveau *un niveau de substitution* ou plus simplement *le niveau correct*. Dit autrement : je peux survivre avec un corps 100% artificiel ou virtuel, c'est-à-dire émulé par un ordinateur. Émulé signifie ici : simulé à un niveau, correct par définition, de substitution.
2. *La thèse de Church*. Une version moderne est que tous les ordinateurs ou systèmes universels peuvent s'émuler les uns les autres. Ce fut l'objet

du chapitre précédent.

3. *Le réalisme arithmétique.* Les propositions de l'arithmétique sont vraies indépendemment de moi. Il s'agit de la croyance en cette réalité mathématique archaïque dont parle si judicieusement Alain Connes[3].

Afin de faciliter la preuve, je vais introduire explicitement quatre hypothèses supplémentaires, que je remplacerai d'un coup par une nouvelle hypothèse. Je montrerai brièvement, en me référant au travail, comment éliminer cette dernière hypothèse supplémentaire. Je procède ainsi pour séparer les difficultés. Les quatre hypothèses supplémentaires sont les suivantes :

1. NIVEAU CORRECT : dans les expériences par la pensée qui vont suivre je vais toujours supposer que les descriptions de corps ou de cerveau ont été effectuées au niveau correct. Ce niveau existe par hypothèse, mais il n'est pas dit qu'une machine puisse scientifiquement déterminer ce niveau correct de substitution. La machine computationnaliste peut cependant *parier* sur le niveau, et nous pouvons raisonner dans le cas où ce niveau a été correctement choisi.
2. UNIVERS CONCRET : je suppose qu'il existe un univers concret, quoi que ce soit précisément. Cette hypothèse donne un décor pour l'argument. Il sera important de voir comment elle sera éliminée. On pourrait parler de "physique de grand-mère" : "concret" signifie pouvant exister de façon singulière, comme sont supposés exister les objets de la vie quotidienne.
3. NEURO : c'est l'hypothèse du neurophysiologiste. Je suppose que le niveau de description de mon cerveau est "assez haut". On verra que le raisonnement ne dépend pas *in fine* du choix du niveau, ni même de ce qu'on entend précisément par cerveau. Le raisonnement ne dépend pas de la question, fort débattue par les philosophes de l'esprit anglo-saxon, de savoir s'il faut inclure l'environnement ou pas, dans ce qu'il est nécessaire de simuler, pour que je survive à la substitution.

[3]Par exemple dans le livre publié aux Éditions Odile Jacob, avec la participation d'André Lichnerowicz et Marcel Paul Schützenberger (2000) : "Triangle de pensées".

La position qui s'oppose le plus au réalisme mathématique est la position conventionnaliste : les propositions mathématiques seraient conventionnelles. Cette position rendrait inintelligible le comportement des mathématiciens qui cachent des résultats mathématiques lorsque ceux-ci ne leur plaisent pas. L'exemple le plus célèbre est celui des Pythagoriciens qui cacheront la preuve de l'irrationalité de la racine de 2. Pourquoi cacherait-on des conventions ? Le réalisme arithmétique est quasi-unaniment accepté par les mathématiciens depuis la nuit des temps. La plupart des critiques philosophiques contre le réalisme mathématique sont le fait de philosophes confondant les théories mathématiques (comportant forcément de nombreux choix conventionnels) avec leur objet.

4. 3-LOCALITÉ : c'est une hypothèse extrêmement faible que je mentionne en raison du rôle clé qu'elle joue dans le raisonnement. Cette hypothèse dit que, par exemple dans notre univers concret, il est possible de séparer deux ordinateurs de telle sorte que leurs calculs n'interférent pas. Ceux qui utilisent un ordinateur font implicitement cette hypothèse. Au cours du raisonnement on comprendra pourquoi je parle de "3-localité".

Le renversement est une conséquence presque directe d'un résultat préliminaire que j'appelle *lemme d'invariance généralisé*. (Je rappelle qu'un lemme est le mot usuel, chez les mathématiciens, désignant un résultat préliminaire). Anticipativement, le lemme d'invariance dit que les expériences subjectives, de la première personne, sont invariantes pour une série de transformations objectives, de la troisième personne. Ces termes vont cependant être définis au fur et à mesure que le raisonnement procède.

Le cas de la télétransportation simple. Avec le computationnalisme et les hypothèses supplémentaires, le *computationnaliste pratiquant* survit à l'usage de la télétransportation (il monte dans le translateur). Il accepte de se faire scanner à Bruxelles, à un certain niveau de description de son corps, de se faire ensuite annihiler (tout ça sous anesthésie par exemple, et dans une cabine que je vais appeler cabine de scanning-annihilation) sachant que l'information numérique obtenue est envoyée à Marseille (pour fixer les idées). A Marseille, à partir de cette information, le candidat est reconstitué, dans une cabine de reconstitution. La survie à la téléportation simple découle du fait qu'il s'agit d'une greffe de corps artificiel déguisée, et que la survie ne dépend que de l'adéquation du niveau de description, pas de la modalité de la reconstruction du corps.

Du point de vue d'un observateur extérieur, je dirai du point de vue d'une tierce ou d'une *troisième personne*, le candidat semble avoir voyagé de Bruxelles à Marseille. Du point de vue du candidat lui-même, je dirai du point de vue de la *première personne*, il semble aussi s'agir d'un voyage de Bruxelles à Marseille. Dans le cas de la téléportation simple la distinction entre première et troisième personne est floue. Ce ne sera déjà plus le cas dans l'expérience suivante.

De la téléportation à l'automultiplication. Considérons le cas de la télétransportation avec délai. A Marseille, cette fois-ci, au lieu de reconstituer le candidat au moment où l'information arrive, on la stocke pour une durée d'un an. Ensuite on procède à la reconstitution. La reconstitution est supposée toujours être faite dans une cabine qui n'a aucun moyen de mesurer

le temps—pas de fenêtre sur l'extérieur, par exemple. Mais la cabine dispose d'un système d'autolocalisation par satellite (du type GPS), et donc le candidat peut savoir s'il est à Marseille. Le candidat peut-il distinguer l'expérience précédente, c'est-à-dire sans délai, de celle avec délai. Avec nos hypothèses, bien sûr que non. Du point de vue du candidat les deux expériences ne sont pas distinguables. Définissons plus précisément le discours de la première personne par les résultats des expériences tels qu'elle les décrit dans un carnet qu'elle transporte (et donc télétransporte) toujours avec elle. Ce compte-rendu personnel sera le même dans la télétransportation simple et dans la télétransportation avec délais, du genre : 'Ok, ça a marché, le GPS confirme que je suis à Marseille, je vais bientôt sortir de la cabine'. Par contre, pour un observateur extérieur (extérieur aux cabines de téléportation) les expériences de téléportation avec délai et sans délai sont très différentes. Avec délai, l'expérience dure un an de plus du point de vue de la troisième personne.

Lemme 1 Retenons (c'est notre premier résultat d'invariance) que les délais de reconstitution ne sont pas 1-observables—ne sont pas observables par la première personne.

A présent nous savons, avec l'hypothèse de 3-localité, que si un candidat survit à une expérience de téléportation, de Paris à Washington, pour prendre un autre exemple, il survit indépendemment de toute activité de calcul ou même en fait de tout événement suffisamment éloigné de la reconstitution.

Considérons alors l'expérience beaucoup plus délicate suivante. Le candidat, après avoir été convenablement scanné à Paris est reconstitué à Washington et simultanément à Moscou. L'information, avec le computationnalisme, est purement numérique, et donc parfaitement duplicable, comme l'amibe. L'hypothèse de 3-localité entraîne que le candidat survit. Mais où ? C'est ici qu'il est important de bien distinguer entre le discours de la troisième personne et les discours des premières personnes possibles. Du point de vue d'un observateur extérieur (troisième personne), le candidat a survécu à Washington *et* à Moscou. Le candidat lui-même, au cas où il a été prévenu de la double reconstitution, peut lui-même dire qu'il sera, après l'expérience, à Washington *et* à Moscou. Mais dans ce cas il parle de lui-même à la troisième personne. Dans son carnet personnel, qui est lui-même dupliqué puisqu'il le transporte avec lui, il devra noter le résultat d'autolocalisation du système GPS de la cabine où il a été reconstitué. Après la duplication, chacune des personnes reconstituée obtiendra un résultat unique et bien précis : soit Washington, soit Moscou. Un carnet personnel contiendra la mention "Je suis reconstitué à Washington" et l'autre "Je suis reconstitué à Moscou". Aucun ne contiendra "je suis reconstitué à Washington et à Moscou". Celui qui se promène à Wash-

ington peut croire *intellectuellement* qu'il a un double reconstitué à Moscou, et réciproquement. Cette connaissance est intellectuelle, communicable à la troisième personne, et non directement accessible, comme la connaissance subjective, privée, de la première personne.

Si on pose alors la question à un candidat pour une telle double-reconstitution : "Où vas tu te sentir être subjectivement après l'expérience", il doit reconnaître qu'il ne pourra pas—toujours avec toutes nos hypothèses—se sentir survivre aux deux endroits simultanément. Et comme il admet survivre à la téléportation et donc (par 3-localité) à la duplication, il doit reconnaître qu'il va se sentir survivre à Washington *ou* à Moscou. Il doit reconnaître qu'il va écrire dans son carnet "Washington" ou qu'il va écrire dans son carnet "Moscou" ; ce qui n'entraîne pas qu'il va écrire dans son carnet "Washington *et* Moscou".

Et ainsi il doit reconnaître qu'à moins de privilégier arbitrairement une reconstitution, il ne peut prédire avec certitude le résultat de la duplication tel qu'il le vit personnellement[4].

Du point de vue de la troisième personne, la situation est parfaitement en accord avec le 3-déterminisme[5] habituellement associé au mécanisme. Mais c'est ce 3-déterminisme qui entraîne la similarité numérique des deux reconstitutions, et c'est cela qui entraîne l'indéterminisme strict du point de vue de la personne reconstituée. En résumé, le 3-déterminisme computationnaliste entraîne un indéterminisme *à la première personne*, naturellement désigné par le terme de 1-indéterminisme.

Lemme 2 Le 3-déterminisme entraîne le 1-indéterminisme.

Un autre fait remarquable est l'existence d'une forme de non localité. De la même façon que le 3-déterminisme entraîne le 1-indéterminisme. La 3-localité entraîne une forme de 1-non-localité. En effet, des événements lointains ne peuvent pas changer la vérité de la survie à une reconstitution, mais l'événement lointain d'une reconstitution identique peut, du point de vue de la première personne, changer son espérance de survie à tel ou tel endroit. Ainsi des événements lointains peuvent changer des prédictions locales du point de vue de la première personne. Par exemple, si un phénomène cosmique lointain devait vous reconstituer dans un état physique *computationnellement* semblable à votre état actuel, vous devriez en tenir compte

[4]Voir la thèse pour plus de détails, voir aussi mon article "Informatique théorique et philosophie de l'esprit" Toulouse 1988.

[5]Dorénavant j'utiliserai des expressions de la forme 1-*quelque chose* ou 3-*quelque chose* pour désigner le quelque chose considéré respectivement du point de vue de la première ou de la troisième personne.

pour prédire votre expérience subjective prochaine. On peut utiliser le fait qu'une absence d'annihilation est équivalente à une annihilation suivie d'une reconstitution immédiate (avec un délai nul) ce que je précise plus bas.

On a :

Lemme 3 La 3-localité entraîne la 1-non-localité.

Considérons à présent une expérience qui mélange la 1-invariance par rapport au délai et le 1-indéterminisme. Le candidat est prévenu à nouveau avant d'entrer dans la cabine de scanning+annihilation qu'il sera reconstitué à Moscou et à Washington, mais un délai d'un an est prévu pour la reconstitution à Washington. Une conséquence du lemme 1, la 1-invariance par rapport au délai, est que du point de vue du sujet ces deux expériences ne sont pas distinguables. Il en résulte que, quelle que soit la façon de quantifier l'indéterminisme dans une expérience d'automultiplication (avec une distribution de probabilité, avec une distribution de masses de croyance, etc.) celle-ci doit être invariante pour l'introduction des délais. Par exemple *si* on quantifie le domaine {Moscou, Washington}, vu comme un ensemble d'-expériences possibles de sa conscience, avec une distribution uniforme de probabilité dans l'expérience de duplication sans délai, *alors* on doit admettre la même distribution uniforme de probabilité sur le domaine {Moscou, Washington} dans l'expérience de duplication avec délai. Les délais entre les reconstitutions ne changent en rien les 1-espérances.

On peut appliquer ce principe dans l'expérience de téléportation *sans* annihilation de l'original. Dans cette expérience vous vous télétransportez de Bruxelles à Lille (par exemple), et comme vous n'êtes pas annihilé à Bruxelles, une tierce personne vous verra à Bruxelles et à Lille. La télétransportation sans annihilation de l'original est équivalente à une duplication. Il s'agit en particulier d'une duplication dont une des composantes a un délai de reconstitution nul (Bruxelles). On admet que ne pas être annihilé est équivalent (avec l'hypothèse du computationnalisme) à être annihilé et reconstitué *sans* délai. Donc si on quantifie {Moscou, Washington} d'une certaine façon dans l'expérience de duplication avec annihilation de l'original, on doit quantifier {Bruxelles, Lille} de la même façon, dans l'expérience de téléportation simple sans annihilation de l'original. Ceci va être utilisé plus tard et donc retenons-le sous la forme du lemme 3bis. C'est le lemme 3 avec une *absence d'annihilation* interprétée explicitement comme annihilation suivie d'une reconstitution sans délai.

1-invariance réelle/virtuelle.

Un dernier principe d'invariance est nécessaire pour conclure avec le déployeur universel. Ce dernier point est difficilement vraiment original : il est lié au vieil argument métaphysique du rêve, que l'on peut trouver chez les logiciens Hindous, les taoïstes chinois, chez Platon, notamment dans le Théétète, chez Descartes, Berkeley, Borges, Lewis Carroll, etc. Roger Caillois a écrit un joli petit livre sur l'argument "L'incertitude qui vient des rêves". Selon cet argument[6] lorsqu'on est éveillé, on ne peut pas tenir *pour sûr* que l'on est éveillé. Avec l'hypothèse computationnaliste on peut remplacer, pour le raisonnement, le rêve par la réalité virtuelle. Dès que l'on simule par ordinateur, avec une précision suffisante, un voisinage d'un environnement, une première personne ne peut distinguer un voisinage réel de ce voisinage virtuel. Cette précision suffisante existe grâce à l'existence d'un niveau de substitution numérique admise par hypothèse. Le remplacement de voisinages réels par des voisinages virtuels ne change donc pas non plus les 1-espérances.

Tous ces principes d'invariance mis ensemble donnent le lemme d'invariance général :

> **Lemme d'invariance général :** La façon de quantifier le 1-indéterminisme dans les expériences d'automultiplication est indépendante des 3-lieux et des 3-moments des reconstitutions, ainsi que de la nature réelle ou virtuelle de ces reconstitutions.

Le renversement. Je vais introduire une nouvelle hypothèse *supplémentaire* qui va remplacer toutes les autres. Je vais toujours supposer qu'il existe un univers concret mais je vais supposer en outre qu'un déployeur universel concret (DUC), y est concrètement et *intégralement* éxécuté. Comme une telle exécution est infinie, cela nécessite que l'univers concret est infiniment extensible dans l'espace et dans le temps, permettant au DUC de ne jamais s'arrêter. Je vais montrer que le renversement est une conséquence directe du computationnalisme *accompagné* de l'hypothèse de l'existence d'un DUC.

Revenons à l'expérience "réelle" du lâcher du crayon sur le pupitre. Je m'apprète à lâcher le crayon. J'aimerais prédire ce qui va se passer. Dans la vie de tous les jours on va en réalité utiliser une "théorie" intuitive du genre "Chaque fois que je lâche un objet il tombe, donc je m'attends à ce qu'il en soit de même : le crayon va tomber sur le pupitre. Une théorie plus sophistiquée est "Mon crayon obéit au lois de la physique, et en l'occurence à la loi selon laquelle tous les corps s'attirent ...". Ici la théorie est plus

[6] J'examine cet argument en détail dans mon rapport technique IRIDIA 1995.

précise et elle permet si on mesure convenablement la position du crayon au départ d'éventuellement décrire précisément la trajectoire du crayon. Avec un DU concrètement et intégralement éxécuté, un computationnaliste doit reconnaître que les théories précédentes sont finalement assez mystérieuses. En effet, au moment où il s'apprète à lâcher le crayon, vu de son point de vue de première personne, il doit reconnaître qu'il effectue une expérience d'automultiplication sans annihilation de l'original. En effet le DU concret va le reconstituer une infinité de fois dans l'état, décrit au bon niveau, où il ressent cette expérience personnelle de s'appréter à lâcher le crayon. En effet, quelque soit le niveau de description nécessaire de l'état permettant à la reconstitution de survivre, le DU va atteindre tôt ou tard (plutôt tard en fait) cet état et générer *toutes* les suites computationnelles possibles. En vertu du lemme d'invariance général, pour prédire son avenir de première personne il doit quantifier l'indéterminisme de la première personne sur l'ensemble de tous ces états virtuels (car émulé par une machine universelle, en l'occurence le DUC) et cela indépendemment du temps et du lieu et du caractère virtuel des reconstitutions.

A priori ce super-indéterminisme entraîné par le DUC est trop fort. Il existe clairement des histoires computationnelles, des rêves de machines comme je les appelle quelques fois, dans lesquelles le crayon, au lieu de tomber, s'élève et se transforme en cochon volant. En effet une telle histoire ou mémoire subjective est une hallucination possible, et en tant que telle, sera générée par le DUC, une infinité de fois aussi. Expliquer pourquoi "je vais fort probablement voir le crayon tomber" revient à justifier la rareté des histoires/mémoires aberrantes, genre cochon volant ou lapin blanc *avec montre et gousset*[7], à partir de la quantification de l'indéterminisme sur l'éxécution du DUC. Cela revient à extraire la physique "correcte" (avec l'hypothèse du computationnalisme) à partir des histoires/mémoires computationnelles possibles générées par le DU. La "physique" est ainsi ramenée à une somme (ou une intégrale) sur tous les calculs possibles, elle est ramenée à la recherche d'une mesure sur les histoires/mémoires computationnelles possibles.

Remarque. On pourrait craindre que le raisonnement précédent ne mène à une forme de solipsisme (la doctrine selon laquelle 'je' serais le seul à rêver). On peut cependant se convaincre qu'en duplicant des populations de machines, l'indéterminisme de la première personne est communicable à la troisième personne *au sein* des populations multipliées. Cela permet d'introduire un indéterminisme de la troisième personne, qui en fait n'est qu'une forme d'indéterminisme à la première personne *du pluriel*. On peut

[7](Re)voir "Alice aux pays des Merveilles".

prédire que si une population de machines partage une histoire computationnelle suffisamment profonde[8] alors si elles observent leur environnement universel le plus probable à un niveau inférieur au niveau où les populations sont multipliées, elles seront confrontées à l'indéterminisme de la première personne du pluriel, ou si vous préférez, aux "univers parallèles", aux *autres* calculs possibles, ou encore aux contrefactuelles. Le fait que l'indéterminisme quantique semble communicable et vérifiable à la troisième personne *et* le fait que l'indéterminisme quantique *pourrait* être un cas particulier de l'indéterminisme computationnaliste rend le solipsisme encore moins plausible. Paradoxalement la confirmation quantique du computationnalisme rend *notre* réalité physique, à la première personne du pluriel, encore plus solide.

Tout se passe comme dans le roman de science-fiction de Daniel Galouye "Simulacron 3", où le héros finit par découvrir la nature virtuelle de son environnement en le scrutant de près. Ici les "bizarreries quantiques", dont l'ordinateur quantique en est un des plus illustres exemples, commencent à qualitativement ressembler à un indice du computationnalisme. Je renvoie à la thèse pour plus de commentaires à ce sujet.
Exercice : montrer que COMP + DUC entraîne une forme d'immortalité. Discuter.

Il reste à éliminer DUC pour terminer la démonstration. Supposons qu'on parvienne à extraire une mesure unique sur les histoires/mémoires computationnelles relatives permettant une quantification précise de l'indéterminisme (et du déterminisme) à partir de laquelle on retrouve les lois de la physique. Dans ce cas, parce que ces lois appartiendront en quelque sorte aux discours nécessaires des machines universelles (autoréférentientiellement correcte, honnêtes) relativement à leurs histoires computationnelles les plus probables, une application élémentaire du rasoir d'Occam et du réalisme arithmétique nous permet de réaliser l'économie de l'hypothèse DUC et UC : nous n'avons pas besoin de postuler l'existence d'un univers concret ni d'un déployeur concrètement exécuté en son sein pour justifier les croyances et les observations probables des machines universelles probables. Le succès de cette solution repose sur le succès de l'extraction qualitative et quantitative de la physique à partir des histoires/mémoires computationnelles relatives. *Personnellement* je pense que l'apparition de l'indéterminisme de la première personne (du pluriel notamment),de formes de non-localité, mais surtout du

[8]Par profond j'entends en gros "issu d'un calcul nécessairement long". On peut rendre cela plus précis avec la notion de profondeur logique de Bennet 1988. On consultera le Rapport Technique IRIDIA 1995.

fait de l'apparition d'une logique quantique à partir de la logique gödelienne des discours possibles des machines universelles, expliqué brièvement au chapitre suivant, et pas mal d'autres faits mentionnés dans la thèse, sont des faits encourageants à cet égard. Le rasoir d'Occam suffit dans ce cas.

D'un point de vue strictement déductif cependant, on peut se passer du rasoir d'Occam et de la confirmation "empirique" par la mécanique quantique, pour éliminer DUC. Il faut alors utiliser l'argument du graphe filmé ou l'argument de Maudlin[9]. Indépendemment Maudlin et moi avons montré l'incompatibilité du matérialisme et du computationnalisme. Comme Maudlin postule le matérialisme, il réfute le computationnalisme. Comme je postule le computationnalisme, je réfute le matérialisme (Marchal 1988, Maudlin 1989). Je renvoie au chapitre 4 de la thèse, ou au rapport technique IRIDIA 1995 ou à l'article de Maudlin 1989. Pour parler franchement je ne suis pas encore satisfait de ma présentation de l'argument du graphe filmé. A certains égards la présentation de Maudlin est meilleure et plus informative. Maudlin semble ignorer la thèse de Church et semble ignorer ainsi la non trivialité a priori des discours des machines sur leurs histoires possibles. Il ne s'est pas non plus aperçu que son argument ne dépendait pas du niveau de substitution. C'est comme ça que je m'explique qu'il passe à coté du renversement.

La "physique" a été ramenée à la recherche d'une mesure sur les histoires/mémoires computationnelles possibles. La démonstration a été construite sur un minimum de psychologie "populaire" sans laquelle, par ailleurs, aucune page de ce livre n'aurait de sens. C'est grâce à cette psychologie populaire que la réduction de la physique à la psychologie n'a pas nécessité de définir exactement ce qu'on entend par les *histoires/mémoires*.

A présent il y a une différence entre montrer que la physique doit être réduite à la psychologie, et montrer *comment* réduire la physique à la psychologie. Je rappelle que la psychologie est (re)définie par les discours autoréférentiellement correctes des machines. Le chapitre suivant suggère une façon de progresser. Il isole une psychologie "exacte", qui élimine (méthodologiquement) la psychologie populaire, et permet *in fine* une extraction de la structure générale des propositions de la physique. C'est *forcément* un peu plus technique. Nous allons d'une certaine façon "interviewer" la machine universelle.

[9]Voir la thèse ou le rapport technique IRIDIA 1995

Chapitre 8

La machine et son ange gardien ("1971" revisité)

Les hommes sont naturellement mus par deux sortes de discours, dont l'un est démonstratif et l'autre, non démonstratif.
Averroès, Commentaire Moyen sur la Poétique[1].

On pourrait se demander si je n'ai pas été inconsistant. Ne suis-je pas en train de communiquer le secret de l'amibe, sous la forme du computationnalisme (je survis à la greffe de cerveau, je survis à la téléportation) ?

La solution intuitive, celle de Ames & Wyler, consiste à passer du secret à la question ou au pari. L'expérience par la pensée de l'autoduplication sans annihilation de l'original permet en effet de se convaincre qu'aucune expérience "scientique" ne peut prouver, c'est-à-dire communiquer à la troisième personne, l'hypothèse computationnaliste, en particulier sous la forme d'une preuve constructive de l'existence d'un niveau adéquat de substitution.

Cela donne cependant un curieux statut à la prémisse : un statut d'interrogation nécessaire ou un statut de prémisse absolu, ou encore d'hypothèse nécessairement hypothétique. Formellement on pourrait craindre bâtir sur un fond mouvant.

Une critique similaire est souvent adressée à tout ceux qui communique sur l'incommunicable, ou parle d'ineffable.

Ainsi, lorsque le jeune (et très positiviste) Wittgenstein énonce dans son "Tractatus" la célèbre formule "Ce dont on ne peut parler, il faut le taire", on est en droit de lui demander de quoi il parle. Et peut-il en parler s'il faut le taire ?

[1]Butterworth C., Harîdiî A. A. (Editeurs), Le Caire, 1987. Traduction citée par Ali Benmakhlouf dans son livre *Averroès*, Les Belles Lettres, Paris, 2000.

De même Lao Tseu ne rate-t-il pas l'occasion de se taire lorsqu'il affirme que le Tao qui a un nom n'est pas le Tao ?

Pour en avoir le cœur net l'idée la plus simple ou la plus naïve, et qui est celle que j'ai entrevu en 1971, est d'interviewer la machine universelle, en modélisant ou capturant, en fait, la communication honnête par la preuve formelle. À l'époque bien sûr—voir les chapitres précédents—je ne connaissais pas la thèse de Church et je n'ai pas saisi de suite la portée du théorème de Gödel pour toutes les machine ; je pensais plutôt à un *machin* comme le PRINCIPIA MATHEMATICA de Russell et Whitehead, ou à PA, c'est-à-dire l'arithmetique formelle de Peano, véritable *Escherichia Coli* des chercheurs en autoréférence[2].

Effectivement, je ne vais pas interviewer n'importe quelle machine universelle. Je vais interviewer celles qui sont autoréférentiellement correctes, en particulier consistantes[3].

Le fait que ces machines soient universelles est, d'un certain point de vue équivalent, au fait qu'elles savent prouver toutes les propositions dites Σ_1 (prononcé sigma 1), pour autant qu'elles soient vraies, bien sûr.

Une proposition est dite Σ_1 si elle est (prouvablement) équivalente à une proposition de la forme $\exists n P(n)$ avec $P(n)$ "mécaniquement", communicablement, vérifiable ou réfutable.

En fait ce que je vais raconter ici fonctionne pour toutes les théories formelles, ou machines, capables de prouver *suffisamment* de théorèmes de l'arithmétique élémentaire. Gödel a découvert que pour de telles machines, il est toujours possible de traduire la proposition "p est prouvable par moi" ou "je sais prouver p" *dans* le langage de la machine. On peut identifier ce

[2]Ce chapitre suppose un minimum de connaissance en logique propositionnelle classique. Consulter par exemple le remarquable petit livre "Introduction à la logique" de François Rivenc, Éditions Payot, Paris, 1989.

Pour la logique modale on peut consulter le livre de Jean-Louis Gardiès, Essai sur la logique des modalités, Presses Universitaires de France, 1979. Un traité classique est le livre de Chellas 1980. On peut évidemment consulter l'annexe modale de la thèse ou le chapitre "Théologie et Modalité" dans le rapport technique IRIDIA 1995.

Notons qu'il existe en anglais une introduction, *récréative*, par Raymond Smullyan, aux logiques modales gödéliennes, de l'autoréférence ou de la prouvabilité, comme on dit aussi : Forever Undecided, 1987, Knopf, New-York, ou en version de poche : 1988, Oxford University Press. Raymond Smullyan est l'auteur d'un très grand nombre de livres soit techniques, soit récréatifs, soit philosophiques qui tous illustrent la profondeur des résultats d'incomplétude de Gödel. Pour la logique de l'autoréférence, ou de la prouvabilité, les classiques sont Boolos 1979 et 1993, ainsi que Smoryński 1985. Rucker 1982 est une autre introduction captivante au théorème d'incomplétude.

[3]Je rappelle qu'une machine ou une théorie formelle est consistante si et seulement si elle ne prouvent pas de propositions fausses.

langage à une portion de l'arithmétique élementaire par l'intermédiaire de codage approprié[4].

"prouvable ⌜p⌝", que je vais noter $\Box p$, peut alors être défini par une formule arithmétique, et même Σ_1 : "il existe n tel que n est une représentation numérique d'une preuve de ⌜p⌝", où ⌜p⌝ désigne une représentation numérique de la proposition p. On dit aussi que ⌜p⌝ est le nombre de Gödel de la proposition p. Et effectivement il est mécaniquement vérifiable ou réfutable qu'un nombre n est le nombre de gödel de la preuve de la proposition p (qui a pour nombre de Gödel ⌜p⌝).

Ces machines sont automatiquement sujettes au lemme de diagonalisation de Gödel[5], dont j'ai déjà parlé au chapitre 3. En particulier il existe des propositions p telles que la machine peut prouver $p \leftrightarrow \neg\Box p$. Il est facile alors de se convaincre que p est automatiquement vrai et *non* prouvable pour la machine consistante. En effet si p, qui est équivalente à $\neg\Box p$, était fausse, $\neg p$ serait vraie, mais $\neg p$ est équivalente à $\Box p$, et donc p serait fausse et prouvable, et la machine serait inconsistante. On voit donc qu'il existe des propositions vraies non prouvables pour les machines universelles consistentes capables de prouver suffisamment de théorèmes de l'arithmétique élémentaire. C'est le premier théorème d'incomplétude de 1931 de Gödel. Le second théorème d'incomplétude, sur lequel je reviens plus loin, affirme que la consistance de la machine, $\neg\Box\bot$ est une telle proposition, vraie et indémontrable par la machine.

Si nous parions que nous sommes de telles machines, nous disposons alors d'un moyen de communiquer sur nos facultés de communicabilité et d'incommunicabilité ; exactement ce que nous recherchions. J'identifie la communicabilité, honnête ou scientifique, de la part d'une machine avec le prédicat de prouvabilité formelle de cette machine. En ce sens, comme dit John Myhill, les théorèmes d'incomplétude de Gödel sont les premiers théorèmes d'une psychologie exacte. Une machine consistante ne peut pas donner une preuve formelle de sa propre consistance[6].

[4]On suppose bien sûr qu'on a fixé un niveau de description de la machine, par exemple celui correspondent au niveau de survie à la substitution.

[5]Ceci est vrai même sans leur qualité d'introspection décrit plus haut.

[6]Notons que le résultat est plus général et concerne aussi les machines ayant accès à des oracles (dans le sens de Turing 1939), c'est-à-dire des données infinies qu'elles peuvent consulter le cas échéant. Vraisemblablement ce chapitre a une portée dépassant le cadre de l'hypothèse computationnaliste. Stricto sensu, le computationnalisme va apparaître dans ce chapitre lorsqu'on va restreindre l'intérprétation des variables propositionnelles (p, ...) aux propositions Σ_1.

On va s'intéresser en outre particulièrement aux machines qui possèdent en plus un minimum de capacités introspectives. Elles savent non seulement prouver toute les propositions vraies Σ_1 mais aussi, elles savent prouver *cela*, dans le sens qu'elles savent prouver, quelle que soit p Σ_1 :

$$p \rightarrow \Box p,$$

où $\Box p$ représente le prédicat interne de prouvabilité. On dit qu'elles savent prouver leur propre Σ_1-complétude. Elles savent ou peuvent savoir qu'elles sont (au moins) universelles[7].

On a vu que le prédicat "prouvable($\ulcorner$p$\urcorner$)", abrévié par $\Box p$, peut être traduit dans le langage de la machine par une proposition Σ_1. Donc, pour ces machines on a, grâce à leur qualité d'introspection, qu'elles savent communiquer, quelle que soit la proposition p du langage de la machine (ou du langage de l'arithmétique) :

$$\Box p \rightarrow \Box\Box p$$

Notons que, comme $\Box p$ représente la proposition *arithmétique* prouvable($\ulcorner$p$\urcorner$), $\Box\Box p$ représente la proposition arithmétique "prouvable ($\ulcorner$ prouvable($\ulcorner$p$\urcorner$)$\urcorner$)".

Le logicien Léon Henkin posera une question très intéressante et tout-à-fait naturelle : qu'en est-il des propositions autoréférentielles p qui, au lieu d'affirmer leur propre non prouvabilité, comme la phrase de Gödel, affirment au contraire leur propre prouvabilité ? De telles propositions existent en vertu du lemme de diagonalisation. A priori de telles propositions peuvent être fausses et non prouvables ou vraies et prouvables. Aucune contradiction n'apparaît, à la différence de la proposition gödélienne qui affirme sa propre non prouvabilité.

En 1955, le logicien hollandais M. H. Löb publie un article avec la solution du problème de Henkin. Aussi étrange que cela puisse paraître les propositions de Henkin, qui affirment leur propre prouvabilité, sont *toujours* vraies et prouvables.

Le théorème de Gödel (l'existence de propositions vraies et non prouvables) repose sur une version du paradoxe d'Épiménide. Gödel remplace la proposition d'Épiménide "je suis fausse" par "je suis non prouvable" exprimée dans le langage de la machine. De même la preuve de Löb repose

[7]Ultérieurement "peut être su" sera pris dans le sens faible de "vrai et communicable", et il faudrait dire seulement que pour chaque propositions p Σ_1 la machine sait communiquer $p \rightarrow \Box p$.

sur un curieux petit paradoxe autoréférentiel. Voici en effet une preuve de l'existence du Père Noël ! (Trouvez l'erreur !). Il s'agit d'un exemple de petit raisonnement en logique propositionnelle. Remarquez la double utilisation de la règle du *modus ponens* MP : $\frac{A\ A\to B}{B}$, qui dit que si on a démontré A et si on a démontré $A \to B$, alors on peut déduire B.

Considérons la phrase ou la proposition P, suivante :

Si cette phrase est vraie alors le Père Noël existe.

"Cette phrase" désigne la phrase P tout entière, il s'agit d'une phrase autoréférentielle.

Je vais d'abord démontrer P. P est une proposition conditionnelle, et la prémisse de P est P elle-même. Pour démontrer une conditionnelle, on suppose sa prémisse et on montre que le reste suit. Supposons que la premisse est vraie, c'est-à-dire supposons que P (cette phrase) soit vraie. Alors P est vraie et automatiquement "P entraîne l'existence du Père Noël" est vraie. Mais alors, avec l'hypothèse P, le Père Noël existe par MP. Donc on a montré que si P est vraie, alors le Père Noël existe. Mais ça, c'est exactement ce que dit P. Donc on a montré, sans hypothèses supplémentaires que P est vraie.

A présent P dit justement que si P est vraie alors le Père Noël existe. Or on vient de démontrer P. Par une nouvelle application de MP, on a que le Père Noël existe !

Où est l'erreur ?

Beaucoup pensent que l'erreur réside dans l'usage d'une proposition autoréférentielle. Mais nous savons que les machines universelles assez riches obéissent aux lemmes de diagonalisation, et que l'autoréférentialité est incontournable. Que se passe-t-il alors ?

Rappelez-vous du théorème de Tarski, au chapitre 3 : on ne sait pas traduire le prédicat de vérité *sur* la machine *dans* le langage de la machine. La phrase de Löb, P, n'est tout simplement pas traductible dans le langage de la machine. Cela résoud le paradoxe, au moins en ce qui concerne le monde des machines.

Gödel démontre son théorème d'incomplétude en remplaçant la vérité par la prouvabilité. De même Löb démontre[8] son théorème, qui résoud la question de Henkin, en remplaçant la vérité par la prouvabilité dans la phrase P. Löb démontre en effet que si une machine prouve $\Box p \to p$, alors elle démontre p. Cela donne la réponse à la question de Léon Henkin, puisque la proposition $p \leftrightarrow \Box p$ entraîne la proposition $\Box p \to p$. Et donc si la machine prouve la première, elle prouve la seconde, et on peut alors appliquer le théorème

[8]Cette preuve, à la différence de celle de Gödel nécessite la capacité supplémentaire d'introspection décrite plus haut.

de Löb. C'est vraiment étonnant : cela ressemble à une forme de "wishfull thinking[9]" ou à la méthode Coué : si je prouve que si j'avais prouvé p j'aurais p, alors j'ai prouvé p. Étrange, mais vrai. Et non seulement cela est vrai, mais on peut montrer que la machine elle-même peut prouver ce résultat. En fait, par un preuve qui miroite le raisonnement du paradoxe du Père Noël, la machine peut démontrer :

$$\Box(\Box p \to p) \to \Box p$$

Cette formule est la formule de Löb. On l'interprète souvent comme une manifestation de modestie de la machine. Elle communique qu'une preuve de p entraîne p, seulement lorsqu'elle prouve p.

A présent, $\neg p$ est équivalent à $p \to \bot$ où $\bot$ sert de proposition fausse générique (vous pouvez partout remplacer $\bot$ par $0 = 1$; de même $\top$ servira de proposition générique vraie, que vous pouvez remplacer par $1 = 1$). En remplaçant p par $\bot$ dans la formule de Löb, on trouve le second théorème d'incomplétude de Gödel comme cas particulier :

$$\Box\neg\Box\bot \to \Box\bot,$$

qu'on peut encore écrire

$$\Diamond\top \to \neg\Box\Diamond\top,$$

où $\Diamond p$ est l'abbréviation usuelle de $\neg\Box\neg p$. De même $\Box p$ est équivalent à $\neg\Diamond\neg p$. Dans ce cas, "je suis consistant" qui est "je ne prouve pas le faux", c'est-à-dire $\neg\Box\bot$, devient encore $\Diamond\top$. Le second théorème de Gödel est prouvable par la machine elle-même et revient à dire "si je suis une machine consistante, alors je ne peux pas démontrer que je suis une machine consistante". Cela montre aussi que l'inconsistance $\Box\bot$ est consistante ! La machine qui ne prouve pas le faux *peut* prouver le faux. Elle peut se tromper ou rêver. On a , au sujet de la machine, $\Diamond\Box\bot$.

On découvre ainsi qu'il y a des propositions vraies sur la machine que la machine peut prouver et d'autres propositions vraies concernant toujours la machine que la machine ne peut pas prouver. Avec le second théorème de Gödel on voit que la machine peut justifier hypothétiquement qu'elle ne peut pas prouver certaines propositions, dont l'importante proposition de consistance de soi, $\Diamond\top$, que nous pouvons interpréter librement, largement,

[9]Prendre ses désirs pour la réalité.

et avec un grain de sel, par "je ne communiquerai pas le faux", "je suis éveillé", "je suis honnête", je suis intelligent, je suis conscient[10].

Appelons "proposition modale" les propositions de la logique propositionnelle élémentaire étendues avec les connecteurs $\Box$ et $\Diamond$. Lorsque les variables propositionnelles sont interprétées par des propositions du langage de la machine et lorsque $\Box p$ est interprétée, comme on l'a fait jusqu'ici, par des proposition du style provable($\ulcorner p \urcorner$), appartenant au langage de la machine, on peut se demander s'il existe une théorie formelle de logique modale capable d'axiomatiser correctement et complètement l'interview de la machine.

Cette question, posée par George Boolos, sera résolue dans l'affirmative par Solovay en 1976. Solovay démontre en effet que le système de logique modale, qu'il appelle G, décrit ci-dessous, axiomatise l'intégralité[11] du discours de la machine auto-référentiellement correcte (ou de la prouvabilité formelle dans les théories assez riches). G est donné par les axiomes et règles suivantes :

AXIOMES	:	$\Box(A \to B) \to (\Box A \to \Box B)$	K
		$\Box A \to \Box\Box A$	4
		$\Box(\Box A \to A) \to \Box A$	L
RÈGLES	:	$\frac{A \;\; A \to B}{B}$	MP
		$\frac{A}{\Box A}$	NEC

NEC désigne la règle d'inférence dite de nécessitation : si j'ai démontré p alors je sais démontrer $\Box p$. K est mis pour Kripke, "4" est le nom consacré (mais assez farfelu) de la formule $\Box A \to \Box\Box A$. L est mis pour Löb, bien sûr.

Mais qu'en est-il de l'"incommunicable mais vrai" ? On sait en effet qu'il y a des propositions vraies sur la machine que la machine ne sait pas prou-

[10]Dostoïesvski aurait défini la conscience ainsi : "La conscience, c'est le pressentiment de la vérité accessible par un homme" cité par Oleg Tabakov, homme de théatre soviétique, lui-même cité par J. P.Thibaudat dans l'article "De l'autre côté du rideau rouge", Libération n2, Juillet 1989. Je n'ai pas trouvé le texte original. En utilisant l'interprétation géométrique de Kripke des formules modales (voir l'annexe sur la logique modale dans la thèse) on peut rendre cette définition (axiomatique et partielle) bien plus parlante. Notons bien qu'ici la conscience, l'honnêteté, etc. ne sont pas *identifiés* avec la consistence. Il est juste suggéré qu'il existe des axiomatiques modales capables de capturer des aspects communs à ces notions. Avec une interprétation encore plus large de $\Diamond p$ par "je suis (sur)vivant", la formule modale correspondant au second théorème de Gödel, exprime alors qu'être vivant, c'est pouvoir mourir. Tout ceci est considérablement développé dans "Conscience et Mécanisme" Rapport Technique IRIDIA 1995.

[11]Seulement au niveau propositionnel. Les logiciens russes ont démontré la *haute* indécidabilité de la logique du premier ordre de l'autoréférence. Ces preuves sont détaillées dans le livre de George Boolos 1993.

ver, comme l'autoconsistance $\Diamond\top$, la consistance de l'inconsistance $\Diamond\Box\bot$, l'"autocorrectitude" $\Box p \to p$, etc.

Solovay offre un cadeau inattendu : l'ensemble des formules propositionnelles modales vraies, prouvables *ou non* (toujours interprétées dans le langage de la machine) est *aussi* complètement axiomatisable. En particulier le système suivant, G*, capture l'ensemble des propositions modales vraies, les communicables et les incommunicables *concernant* la machine :

AXIOMES	:	tous les théorèmes de G,	
		$\Box A \to A$	T
RÈGLES	:	$\frac{A \quad A \to B}{B}$	MP

Remarquons la perte de la règle de nécessitation. C'est un exercice facile de montrer que G* + la règle de nécessitation donne un système inconsistant.

Solovay démontre l'adéquation et la complétude de G et G* pour la preuve sur-et-par la machine et le vrai sur la machine respectivement, mais il démontre encore la décidabilité de ces systèmes : G et G*, à l'instar de la logique propositionnelle peuvent être générés[12] "mécaniquement". G* étend G. La couronne G* \ G, devient un système décidable, fermé pour le modus ponens (voir la thèse) capturant les incommunicables vérités propositionnelles de la machine, l'espace infini des secrets de l'amibe.

G axiomatise fidèlement et complètement le discours de la machine consistante (ou honnête, ou éveillée). J'identifie alors parfois G avec le discours de *la machine elle-même*, il suffit de se rappeler de la façon dont les symboles sont interprétés dans le langage de la machine. De la même façon, en l'honneur de Judson Webb (j'explique pourquoi au chapitre 6) j'appelle G* le discours de *l'ange gardien de la machine*. G* ne parle pas de lui, mais parle sur G, ou sur la machine. G* axiomatise la part, aussi bien communicable (G* étend G) que non communicable des vérités sur la machine. Nous pouvons donc interviewer la machine universelle mais aussi son ange gardien.

La troisième personne. Si on modélise, ou même si on capture la communication honnête des machines consistantes par la prouvabilité formelle, on ne sort pas du discours scientifique à la troisième personne. Il s'agit bien d'un discours autoréférentiel, et lorsque la machine communique $\Box p$, elle est bien en train de prouver correctement qu'elle peut prouver p, mais cette autoréférence est une autoréférence à la troisième personne. Lorsque la machine

[12]Voir mon rapport technique IRIDIA pour des démonstrateurs de théorèmes (en LISP) pour G et G* et les autres logiques de ce chapitre, et bien d'autres considération sur la logique de l'autoréférence.

communique $\Box p$, elle communique une proposition arithmétique (par exemple) ou une proposition dans son langage de machine qui est équivalente à "il existe un nombre (une liste, une suite de signes) qui code une démonstration d'une proposition codée par $\ulcorner p \urcorner$". *Nous* savons, éventuellement, si la machine n'est pas trop compliquée, qu'elle est correcte et autoréférentiellement correcte. Mais l'autoréférence gödélienne est quasi-accidentelle dans le chef de la machine. Clairement, ce code, étant une description de soi à un niveau formel, vaut tout autant pour le double de la machine, comme son éventuel doppelgänger après une expérience d'autoduplication.

Dont acte. Et tant mieux, cela donne, *par construction* un discours honnête, (scientifique), et cela garanti que l'entièreté de notre conversation, aussi *philosophique* qu'elle pourra sembler, admet une interprétation en terme de propositions arithmétiques vraies *sur la machine* et prouvables *par la machine*, et éventuellement vraies *sur la machine* et *non* prouvables *par la machine* quand on interview l'ange gardien. Celui-ci parle aussi à la troisième personne, et son discours est aussi scientifique, bien qu'il ne peut être qu'interrogatif de la part de la machine.

Si l'ange gardien peut communiquer des propositions non communicables, c'est parce que celles-ci sont non communicables *par* la machine au sujet de laquelle il parle. G*, à la différence de G, ne parle pas de lui, il parle au sujet de la machine. L'ange gardien peut dire $\Diamond \top$, cela ne signifie pas "*je* suis consistant", cela signifie que *la machine* (*la machine qu'il garde*, si vous voulez) est consistante. C'est le décalage entre la machine et son ange gardien qui éclaire de façon lumineuse, dans le monde des machines, les possibles discours sur l'incommunicable.

Toutes ces remarques s'étendent naturellement aux propositions de la psychologie ou de la physique, ces termes admettant la nouvelle interprétation imposée par le renversement. Il faudra cependant parvenir à traduire les termes de la psychologie et de la physique des machines en terme de communicabilité formelle par la machine ou par son ange gardien.

Dit autrement, pour capturer la ou les logiques de la première personne, il faut, lors de l'interview de la machine universelle et de son ange gardien, traduire le "je" du sujet qui sait, qui mesure ou qui observe, en terme de la prouvabilité formelle à la troisième personne. Ce que je propose d'aborder maintenant.

La première personne qui sait. Le sujet de la première personne est celui qui sait. Il est le sujet de la connaissance. Des formules modales typiques axiomatisant la *connaissance* sont la formule dite de réflexion $\Box p \rightarrow p$ dans des logiques fermées pour la règle de nécessitation $\frac{p}{\Box p}$ (en particulier on veut

$\Box(\Box p \rightarrow p)$. Elles garantissent d'une certaine façon le rattachement ombilical du sujet à la vérité, et ceci à la base. Notons qu'on se restreint ici encore à une connaissance concernant des propositions communicables. On demande que la connaissance de p entraîne la communicabilité de p.

Gödel, dans son petit papier de 1933, avait déjà constaté l'inadéquation du prédicat formel de prouvabilité pour capturer la connaissance[13], puisque $\Box\bot \rightarrow \bot$ est incommunicable, quoique vrai, et bien sûr, pour cette raison $\Box(\Box\bot \rightarrow \bot)$ est faux. En fait l'ange gardien nous a prévenu, bien qu'il affirme que la machine est honnête ($\Diamond\top$) elle peut (il est consistant de) communiquer du faux ($\Diamond\Box\bot$). On peut donc interprété aussi la prouvabilité comme une *croyance* car croire le faux est possible à la différence du savoir. Par exemple, on peut dire "Dominique *croyait* que la terre était plate" ; on ne dira jamais "Dominique *savait* que la terre était plate". Le savoir est donc connecté à la vérité par définition. Et en identifiant la preuve formellement communicable avec la prouvabilité $\Box$ de la machine universelle, on met clairement les propositions de la science et des discours à la troisième personne du coté des propositions crues ... et *accidentellement* sues. L'ange gardien sait que la modestie du discours scientifique de la machine est logiquement attaché à la possibilité de l'erreur ou du mensonge ou du rêve (habituel, non lucide). La machine ne le croit pas, mais elle peut l'inférer, et se mettre à douter.

Si la prouvabilité formelle ne vérifie pas les axiomes naturels de la connaissance, on doit donc essayer de définir la connaissance à partir de la prouvabilité formelle.

La façon la plus simple d'attacher la prouvabilité au cordon ombilical de la vérité serait de définir "p est connaissable" par "p est prouvable *et* p est vraie". En remplaçant "prouvable" par "opinion" ou "opinion justifiable", on retrouve des essais de définition de la connaissance que Théétète propose à Socrate dans le Théétète de Platon[14]. Mais cette définition n'est pas exprimable dans le langage de la machine. C'est à nouveau une conséquence du théorème de Tarski, on ne sait pas définir "p est vraie" *pour* la machine *dans* le langage de la machine.

Cela semble impossible. De façon très générale, il est effectivement facile de montrer qu'il est impossible de définir, pour une machine consistante,

[13]Ceci avait d'ailleurs été déjà développé par Kolmogorov. A. Kolmogorov, 1932, N., Zur Deutung der Intuitionistischen Logik, Math. Zeitschr., 35, pp. 58-65.

[14]Et encore beaucoup discutée actuellement, voir par exemple Burnyeat 1991 dans le beau recueil de Monique Canto-Sperber. Burnyeat M., 1991, Socrate et le jury : de quelques aspects paradoxaux de la distinction platonicienne entre connaissance et opinion vraie, dans Canto-Sperber M. (ed.), 1991, Les paradoxes de la connaissance, essais sur le Ménon de Platon, Éditions Odile Jacob, Paris, pp. 237-251.

un prédicat arithmétisable (ou définissable dans le langage de la machine), et donc sujet au lemme de diagonalisation (voir chapitre 3), vérifiant à la fois la formule de réflexion, $\Box p \rightarrow p$ et la nécessitation $\frac{A}{\Box A}$. En effet—par conséquence directe du lemme de diagonalisation—la machine, pour une certaine proposition k, prouverait $k \leftrightarrow \neg\Box k$. En particulier elle prouverait $\Box k \rightarrow \neg k$. Mais elle prouve la réflexion $\Box p \rightarrow p$ pour toute proposition p, donc elle prouve en particulier $\Box k \rightarrow k$. Par calcul propositionnel elle prouve alors $\Box k \rightarrow (k \,\&\, \neg k)$, c'est-à-dire $\Box k \rightarrow \bot$, c'est-à-dire encore $\neg\Box k$. Comme elle a prouvé $k \leftrightarrow \neg\Box k$, elle prouve k, et par nécessitation, elle prouve $\Box k$. Donc elle prouve $\neg\Box k$ et $\Box k$. Et ainsi elle est inconsistante.

Pour nos machines qui ont assez d'introspection, et qui donc vérifient le théorème de Löb, et même le prouvent, on voit encore plus rapidemment que la réflexion combinée à la nécessitation sont interdites car l'application de la nécessitation sur la réflexion (avec p substitué par $\bot$) donne $\Box(\Box\bot \rightarrow \bot)$, ce qui par Löb, et MP, donne $\Box\bot$, ce qui par une nouvelle application de la réflexion donne $\bot$.

De même que la preuve de Tarski montre que la vérité sur la machine n'est pas exprimable ou définissable dans le langage de la machine, le petit raisonnement ici montre que la connaissance, dans le sens très général de n'importe quoi axiomatisé par la réflexion et fermé pour la nécessitation, n'est pas non plus exprimable (arithmétisable) dans le langage de la machine.

Remarquons bien que G et G* sont cohérent à cet égard : G est fermé pour la nécessitation mais ne prouve pas la réflexion, et G* prouve la réflexion mais ne vérifie pas la nécessitation (forcément parce que s'il obéissait à la nécessitation, cela entraînerait que G, la machine elle-même, prouverait tout ce que lui, l'ange gardien, sait prouver. La couronne serait vide et *l'amibe* ou la machine universelle n'aurait pas de secrets. Mais ceci montre aussi que ni G, ni G* ne capturent *directement* une description du "connaisseur" ou de la première personne).

A des variations techniques près l'argument ici, est souvent présenté comme un moyen d'utiliser le théorème de Gödel pour distinguer l'homme de la machine. En simplifiant : la connaissance n'est pas arithmétisable, c'est-à-dire traductible dans le langage d'une machine, donc les machines n'ont pas de prédicat de connaissance, donc elle ne peuvent pas connaître[15].

En réalité l'argument montre seulement que la connaissance ne peut pas être arithmétisée, qu'elle soit le fait d'une machine ou de quoi ou qui que ce

[15]Ceci est aussi lié au paradoxe du connaisseur de Kaplan et Montague 1961, et au paradoxe de la connaissance chez Platon, bien sûr (cf Monique Canto-Sperber. Voir la référence de Burnyeat 1991).

soit.

Mais comment alors interviewer la machine et l'ange gardien sur la connaissance si on ne sait pas traduire la connaissance dans le langage de la machine ?

Eh bien, une façon très simple[16] est la suivante. Au lieu de définir la connaissabilité de p par "p est prouvable et p est vrai", on va, toujours en s'inspirant de Théétète, définir la connaissabilité de p par "p est prouvable et p". Cela nous permet d'éviter l'usage imposssible, comme on vient de le voir, d'un prédicat de vérité ou de connaissance. On va simplement, au niveau de la logique propositionnelle, définir un nouveau connecteur modal $\boxdot$, où $\boxdot p$ est directement interprété par $\Box p$ & p. On remplace l'usage impossible—par Tarski—de VRAI($\ulcorner p \urcorner$) par la simple assertion de p.

Remarquons que la réflexion de cette connaissabilité est non seulement vraie sur la machine mais est prouvable par la machine : la machine prouve $\boxdot p \rightarrow p$, puisqu'elle prouve évidemment $(\Box p \ \& \ p) \ \rightarrow p$.

De même la logique du discours de la machine sur $\boxdot p$ est fermée pour la nécessitation. En effet avec la nécessitation $\frac{p}{\Box p}$ et la règle $\frac{p}{p}$ (qui dérive elle-même de $p \rightarrow p$ avec le modus ponens), on a que si la machine prouve p elle prouve p & $\Box p$, donc elle est fermée pour $\frac{p}{p \ \& \ \Box p}$, c'est-à-dire $\frac{p}{\boxdot p}$.

Il n'y a pas de paradoxe : $\boxdot$ n'est *pas* arithmétisable. De même que le théorème de Tarski montre qu'il n'y a pas de prédicat de vérité définissable dans le langage de la machine, ce qui veut dire qu'il n'y a pas de $V(x)$ tel que la machine prouve $V(\ulcorner p \urcorner) \leftrightarrow p$, il n'y a pas de prédicat de connaissance *théétètique* définissable dans le langage de la machine, ce qui veut dire qu'il n'y a pas de $C(x)$ tel que la machine prouve $C(\ulcorner p \ \urcorner) \leftrightarrow (p$ & prouvable $\ulcorner p \urcorner)$. La connaissabilité, comme la vérité arithmétique, ne sont pas sujettes à la diagonalisation parce qu'elles ne sont pas définissables en terme purement arithmétique. En définissant la vérité de p par l'assertion pure de p, on découvre un discours non trivial de la machine sur la connaissabilité, qui se passe en quelque sorte de la définition arithmétique et qui se passe de la représentation (du nombre de Gödel) des formules dont on parle.

Boolos, Goldblatt, ainsi que Kuznetsov & Muravitskii fin des années 1970 montreront indépendemment que cette logique de la connaissabilité est complètement axiomatisée, de façon correcte (sound) et complète par un système inventé, en 1967, par un logicien polonais, Grzegorczyk. Il s'agit de la logique standard de la connaissance, la connaissabilité en fait, S4, (K,T,4,

[16]Découverte et étudiée indépendemment par plusieurs logiciens, l'américain George Boolos 1980, le néozélandais Robert Goldblatt 1978, et les Russes Kuznetsov et Muravitsky 1977, dans un contexte plus mathématique éloigné des *paradoxes* de la connaissance cependant. Artemov 1990 propose comme thèse l'équivalence de la prouvabilité intuitive ou informelle avec $\Box p$ & p. Voir le rapport technique de l'IRIDIA 1995.

+ les règles MP et Nec) auquel est ajoutée la formule un peu curieuse de Grzegorczyk Grz :

AXIOMES	:	$\Box(A \to B) \to (\Box A \to \Box B)$	K
		$\Box A \to A$	T
		$\Box A \to \Box\Box A$	4
		$\Box(\Box(A \to \Box A) \to A) \to A$	Grz
RÈGLES	:	$\frac{A\ ,\ A \to B}{B}$	MP
		$\frac{A}{\Box A}$	NEC

Géométriquement, avec la sémantique de Kripke[17], on peut voir cette logique comme une sorte de logique temporelle décrivant des évolutions futures d'états de connaissance se développant irréversiblement (antisymétriquement). Grâce à une suggestion de Gödel[18] selon laquelle S4 peut émuler par l'intermédiaire d'une transformation modale (voir la thèse) la logique intuitionniste (la logique du sujet de Brouwer, formalisée selon Heyting), Kripke découvrira sa (bien connue) sémantique de la logique intuitionniste. Celle-ci décrit ces mêmes évolutions temporelles d'état de connaissance. La logique du sujet ici est une logique du temps "subjectif".

On peut démontrer que le discours de l'ange gardien sur la connaissabilité n'apporte rien de plus que le discours de la machine elle-même. S4Grz*, l'ensemble des propositions modales vraies, et donc prouvable par G* (grâce au théorème de Solovay) est égale à S4Grz. Du point de vue de la connaissabilité (arithmétique) la vérité est équivalente à la prouvabilité. Voir Boolos 1993.

Ce sont ces résultats et ces transformations qui permettront à Goldblatt d'extraire une interprétation purement arithmétique, et donc interprétable dans le langage de la machine, de la logique intuitionniste LI. Si j'énonce ce résultat, ce n'est pas parce qu'il est très intéressant, mais parce que je vais m'inspirer de cette façon de procéder pour questionner G et G* sur le déployeur universel et l'origine logique des croyances dans les propositions de la physique. Notons qu'ici aussi LI = LI* : du point de vue du sujet intuitionniste la vérité est équivalente à la simple assertion.

On peut aussi interroger G, la *machine elle-même* et G*, l'ange gardien, sur des propositions *mixtes*. En particulier, pour toute proposition p (dans le langage de la machine) G* prouve $\Box p \leftrightarrow \boxdot p$, mais la machine ne le prouve pas[19]. Cela illustre que la distinction entre la première personne et la

[17]Voir l'annexe sur la logique modale.

[18]Dans son papier de 1933. Cette suggestion sera prouvée McKinsey et Tarski 1948.

[19]Pour une utilisation de ce fait et fait de ce genre pour une réflexion sur le rêve et l'éveil, voir Conscience et mécanisme, Rapport technique Iridia 1995. J'y montre que G et

troisième personne est une nuance intensionnelle (modale) de la prouvabilité. Il s'agit de points de vue différents de et sur la *même* machine.

La première personne qui mesure, ou qui sent.
Il y a toute sorte de choses remarquables sur ces nuances intensionnelles qui ne dépendent pas du fait qu'elles soient vaccinées contre la diagonalisation, comme $\boxdot$.

En effet en substituant la vérité "assertorique" d'une proposition p, par la possibilité de la vérité, c'est-à-dire en passant de $\Box p \,\&\, p$ à $\Box p \,\&\, \Diamond p$, on définit une nouvelle variante intensionnelle qui est arithmétisable.

Il s'agit d'un raffinement *arithmétique* de l'idée de Théétète[20].

Cela revient à définir un nouveau connecteur modal, $\triangle$ avec $\triangle p$ équivalent à $\Box p \,\&\, \Diamond p$.

Qu'il s'agisse d'une variante intensionnelle est assuré par G*. Ici aussi G* prouve $\triangle p \leftrightarrow \Box p$, mais G ne le prouve pas. De même G* prouve $\triangle p \leftrightarrow \boxdot p$ mais G ne le prouve pas. Du point de vue de l'ange gardien il s'agit toujours de la même machine (définie à la troisième personne par les propositions qu'elle communique). Du point de vue de la machine il s'agit de nuances distinguant justement différentes sortes de point de vue. La possibilité de ces nuances ont leur source dans le phénomène d'incomplétude.

"$\triangle$" est clairement arithmétisable, c'est-à-dire encore définissable dans le langage de la machine puisqu'il correspond à prouvable($\ulcorner p \urcorner$) & consistent($\ulcorner p \urcorner$).

Donc le prédicat représenté par $\triangle$ est d'office diagonalisable, dans le sens de sujet au lemme de diagonalisation. On peut montrer que la formule "gödélienne" k prouvablement équivalente à $\neg\triangle k$ est équivalente à $\Box\bot \vee \Diamond\Diamond\top$ où "$\vee$" est mis pour le "ou" logique usuel[21].

L'interview de G, la machine elle-même, au sujet des propositions modales utilisant ce nouveau connecteur, donne une nouvelle logique modale, que j'appelle Z. De même l'interview de l'ange gardien G* produit une nouvelle logique Z*. A la différence du système de Grzegorczyk S4Grz, les systèmes modaux Z et Z* sont distincts : $\Box\top$ est prouvable par Z* mais pas par Z,

G* assurent la cohérence de l'argument métaphysique du rêve (ou de la réalité virtuelle) utilisée par Théétète, et son utilisation dans le contexte du computationnalisme. On trouvera aussi un raffinement de l'analyse du cogito de Slezak, ainsi que des variations sur des réfutations de Lucas.

[20]Une des réfutations de Lucas par Judson Webb utilise une idée similaire.

[21]Avec cette nouvelle forme du théorème de Gödel on obtient d'office des raffinements intensionnels du cogito de Descartes du à Slezak, et dans mon rapport technique j'illustre comment utiliser ce raffinement pour critiquer l'argument du philosophe positiviste Norman Malcolm (Oxford) contre l'existence de l'expérience de la conscience à la fois dans les rêves et chez les machines.

puisque $\triangle\top$ appartient à G* \ G. En particulier on voit que la logique Z ne peut pas être fermée pour la règle de nécessitation (comme G*, mais à la différence de G et de S4Grz). Ceci interdit d'utiliser la sémantique de Kripke pour tenter d'axiomatiser Z, et en particulier la question même de savoir si Z et Z* sont complètement axiomatisables reste ouverte. Néamoins il n'est pas difficile d'uliser les travaux de Solovay pour montrer que Z et Z* sont décidables et mathématiquement bien définis[22].

Le raffinement de l'idée de Théétète, qui correspond au passage de "$\Box p \,\&\, p$" à "$\Box p \,\&\, \Diamond p$" modélise le passage de l'actualité de p à la possibilité de p. Ce passage est rendu quasi-obligatoire par l'argument du déployeur universel, où, je rappelle, le computationnaliste qui veut prédire son avenir est obligé de quantifier un indéterminisme de la première personne sur l'ensemble de *toutes* ses extensions *consistantes*. Comme dans les approches philosophiques de l'actuel par l'indexical, où l'actualité est définie par chaque possibilité *vue de l'intérieur"*, le computationnalisme force d'intérioriser la consistance. De plus, avec un nouveau connecteur $\diamondsuit$ mis pour $\neg\triangle\neg$, on peut montrer que G prouve

$$\triangle p \rightarrow \diamondsuit p$$

C'est-à-dire que Z prouve $\Box p \rightarrow \Diamond p$. Cette formule possède un nom standard chez les logiciens modaux : D. "D" est mis pour déontique. Dans les logiques déontiques, l'interprétation du carré modal $\Box$ est l'obligation et la duale $\Diamond$, c'est-à-dire $\neg\Box\neg$ correspond à la permission (p est permis si et seulement s'il n'est pas obligatoire d'avoir $\neg p$, de même que p est obligatoire si et seulement s'il n'est pas permis d'avoir $\neg p$). La formule D indique qu'il faut permettre ce qui est obligatoire, ou qu'il ne faut pas rendre obligatoire ce qui est interdit. C'est un axiome élémentaire du droit.

La formule D est aussi la bienvenue pour la recherche d'une quantification sur un indéterminisme. Dans des systèmes fermés pour la règle de nécessitation D a été utilisé pour capturer la modalité de la certitude en probabilité, et sans nécessitation elle a été utilisée pour modéliser ou capturer des notions de crédibilité. En effet si p est une proposition certaine on

[22] On peut montrer de plus (voir la thèse) que Z admet une sémantique dite des voisinages (notion attribuée aux logiciens Scott et Montague). En logique modale on définit souvent l'intension d'une proposition par l'ensemble des mondes possibles où cette proposition est vraie. La structure logique des propositions de Z confère une structure de quasi-filtre aux voisinages. Un quasi-filtre est un filtre sans élément maximal. Dès qu'on a un filtre on peut construire une sémantique de Kripke. En ce sens Z a une quasi-sémantique de Kripke qui permet d'utiliser une portion non négligeable de l'intuition de Kripke. Voir le livre de Chellas (référence dans la thèse) pour la sémantique de Scott et Montague. Voir le Rapp. Techn. IRIDIA, 1995.

aimerait que $\neg p$ ne soit pas certaine. Cela fait de la version arithmétique de la définition de Théétète, une variante de la prouvabilité gödélienne saisissant une sorte de croyance anticipant son autoconsistence[23].

On peut voir cela d'une autre façon. En terme de mondes possibles, autrement dit avec la sémantique de Kripke, une proposition de la forme $\Diamond p$ est vraie dans un monde M_1 (ou un état ou une situation), ...) si je peux accéder à partir de M_1 à un monde M_2 où p est vraie. Imaginer que vous aboutissiez à une situation qui n'admet aucun échapatoire, un monde duquel vous ne pouvez accéder à aucun autre monde possible, une sorte de cul-de-sac, un dernier monde. A un tel endroit toutes les propositions de la forme $\Diamond\#$ sont fausses, et donc toutes les propositions de la forme $\Box\#$ sont vraies. Dans un dernier monde rien n'est possible et tout est nécessaire ! En attachant la consistance à la prouvabilité, comme dans la version arithmétisable de l'idée de Théétète, on filtre essentiellement les derniers mondes, c'est ce qu'il faut faire puisque les probabilités (ou crédibilités) qui apparaissent dans le renversement sont définies sur les extensions consistantes. Cette justification doit être nuancée vu que, dans notre contexte arithmétique (de machines), la version arithmétique de l'idée de Théétète nous fait perdre la sémantique de Kripke, et notamment la possibilité de structurer les mondes avec des relations d'accessibilité. Mais cette justification peut être corrigée pour fonctionner avec la sémantique de Scott et Montague. On peut voir dans cette idée une sorte de Darwinisme arithmétique : on interroge la machine sur ses extensions ex-

[23]Avec la psychologie folklorique, on peut admettre que "survivre" nécessite de rester conscient. Si on regarde la conscience comme une fille logique de la consistance, par incomplétude elle ne peut pas être *purement* logique. On peut alors voir la conscience, de la machine autoréférentiellement correcte, comme le fruit d'une anticipation instinctive (programmée ou sélectionnée, à un niveau ou à un autre) de la consistance de soi. La décidabilité de G* \ G illustre le caractère inférable d'une collection non négligeable de propositions non prouvables. On n'est pas très loin ici de l'idée de Helmholtz selon laquelle la perception est produite par inférence instinctive. Cela génère une sorte de voyage de G à G*.

La machine peut inférer inductivement une proposition de G* \ G et la conserver comme secret ou paris ou simplement l'affubler d'un point d'interrogation. Mais elle peut aussi intégrer la nouvelle proposition en modifiant son code (à un niveau ou à un autre). Dans ce cas elle se change elle-même, en tant que communicatrice de vérité. G et G* s'appliquent *d'office* à la nouvelle machine, bien que *l'interprétation arithmétique* de G et de G* change puisqu'ils s'appliquent à la nouvelle machine avec son nouveau langage. Cette intégration suggère un rôle pour la conscience : celui de permettre l'accélération relative d'une machine universelle relativement à une autre. En effet, inférer et intégrer une proposition (relativement) consistante, rend décidable une infinité de propositions indécidables, mais aussi raccourcit la longueur d'une infinité de propositions démontrables. Je fais allusion ici au théorème de l'accélération (speed-up) de Gödel. La conscience *de soi* se développe lorsque la non communicabilité de cette conscience/consistance est elle-même anticipée. Cela permet la distinction et la reconnaissance de soi et de l'autre.

clusivement consistantes. On peut aussi y voir une généralisation du principe anthropique, une sorte de principe "machine universelle-tropique" : on interroge la machine consistante sur ses possibles environnements où elle reste consistante. On interdit simplement de quantifier l'indéterminisme sur les "derniers mondes". *In fine* le filtrage sur les extensions consistantes justifiera la logique quantique comme logique des propositions dont une forme de consistance est observable ($\Box\Diamond p$) pour p accessible par le déployeur universelle.

Une autre et dernière motivation d'ordre générale pour le passage de S4Grz à Z et Z* est la suivante. Nous voudrions limiter le nombre d'extensions qui, quoique consistantes, sont aberrantes ; comme les expériences "hallucinatoires" genre cochon volant ou lapin blanc (avec veste et gousset).

Comme en géométrie algébrique, où rajouter des équations à un système d'équations restreint l'ensemble des objets géométriques satisfaisant le système, en logique formelle, rajouter des axiomes restraint la classe de ses modèles[24]. Malheureusement rajouter des axiomes à une théorie assez riche et sujette à la diagonalisation ne diminuent pas vraiment le nombre de modèles à cause de l'infinité des propositions indécidables : il faudrait inclure une infinité de formules si on voulait se débarrasser définitivement du cochon volant de cette façon. Une meilleure idée consiste à affaiblir la logique en espérant multiplier suffisamment les modèles non aberrants. De fait, Z affaiblit assez considérablement S4Grz. De cette façon on augmente le nombre de modèles, et on peut argumenter que l'augmentation obtenue avec Z (et Z*) produit des voisinages ayant la puissance du continu. Il ne reste plus alors qu'à isoler une relation de proximités sur les extensions et à montrer que (presque) toutes nos extensions sont relativement "normales".

L'application de l'idée de Théétète à la logique de l'autoréférence, c'est-à-dire la passage de $\Box p$ à $\Box p \,\&\, p$ définit déjà un espace du connaissable qu'on peut voir comme une réalité psychologique de la première personne.

La version arithmétique de cette idée, c'est-à-dire le passage de $\Box p \,\&\, p$ à $\Box p \,\&\, \Diamond p$, définit une réalité psychologique plus tangible. Il est facile de montrer que ni Z, ni Z* ne prouvent la formule 4 ($\Box p \rightarrow \Box\Box p$), et on peut argumenter que cela fait de cette réalité une sorte de croyance immédiate,

[24]Un modèle est une structure mathématique qui satisfait (rend vrai) une théorie, vue comme un ensemble d'axiomes et de règles. Le logicien, comme le peintre, utilise le mot modèle pour désigner une réalité possible. La théorie, comme le tableau, vise à approximer ou capturer des aspects de cette réalité. Les physiciens utilisent souvent le mot "modèle" pour la théorie ou l'approximation théorique, comme lorqu'on parle de modèle réduit ou de modélisation (par exemple le modèle de Bohr de l'atome). Ceci explique peut-être les fréquents dialogues de sourd entre logiciens et physiciens.

non directement accessible par introspection.

Notre but cependant est d'isoler la physique, ou au moins le squelette ou la structure logique des propositions "observables". L'argument du déployeur universel nous à déjà motivé pour traduire l'observation certaine, vue comme une sorte de croyance immédiate, par la prouvabilité accompagnée explicitement par la consistance. Mais nous n'avons toujours pas introduit le déployeur universel explicitement dans notre interview de G et de G*.

Avec le computationnalisme, l'indéterminisme *physique* n'est pas défini sur toutes nos extensions consistantes, seulement sur celles qui étendent des états atteints par le déployeur universel. On se rappelle qu'un déployeur universel n'est qu'une machine universelle écrasée : un catalogue des histoires générables et des états accessibles. Arithmétiquement, l'universalité peut être modélisée par la Σ_1-complétude, et le déployeur universel peut être considéré comme un catalogue des preuves des propositions (vraies) et Σ_1. De telle propositions sont vérifiables quand elles sont vraies, mais pas nécessairement réfutables lorsqu'elles sont fausses[25].

Nous devons donc limiter l'interprétation arithmétique des variables propositionnelles p aux propositions Σ_1. Nous savons (voir plus haut) que les propositions $p \rightarrow \Box p$ sont vraies pour ces propositions et même prouvables par la machine suffisamment introspectrice[26].

En résumé on obtient la phénoménologie de la matière, en effectuant :

- la Σ_1 restriction,
- la version arithmétique de l'idée de Théétète.

Le résultat donne deux logiques décidables que j'appelle Z_1 et Z_1* et qui correspondent naturellement aux interviews de G et G*. La couronne Z_1* \ Z_1 est non vide. En particulier j'ai pu montrer que Z_1* prouve la formule :

$$p \rightarrow \Box\Diamond p$$

,

avec p représentant exclusivement des propositions arithmétiques Σ_1 (et donc ici il faut restreindre les règles de substitution pour les logiques $Z_1(*)$). C'est le théorème 14 de la thèse.

[25]Notons que Abramsky 1987 et Vickers 1989 ont déjà modélisé la notion d'observable par des propositions similaires.

[26]En fait le système V constitué des axiomes de G, accompagné de $p \rightarrow \Box p$, avec les règles MP et NEC, est non seulement correct, mais a été prouvé arithmétiquement complet pour la (Σ_1) prouvabilité. De même le système *naturel* V* est complet pour la vérité sur ces propositions (Visser 1985).

On peut dire qu'il y a, concernant notre recherche d'une phénoménologie purement arithmétique de la matière, une bonne nouvelle et une mauvaise nouvelle.

- La bonne nouvelle : $p \rightarrow \Box\Diamond p$ est une formule modale, appelée B dans la littérature, permettant d'axiomatiser la logique des propositions de la mécanique quantique. Ceci vient d'un résultat obtenu par Robert Goldblatt (Goldblatt 1974). Comme S4Grz, qui formalise à une transformation modale près, la logique intutionniste du sujet, le système B, axiomatisé par les axiomes K, T, B avec les règles MP et NEC, formalise, à une transformation modale près, la logique quantique. Un miracle s'est opéré ici, car on est parti d'une logique de la connaissance antisymétrique pour arriver à une logique quasi-symétrique, comme la formule B le suggère.
- La mauvaise nouvelle est que Z_1*, comme G* par ailleurs, n'est pas fermée pour NEC, ni même, à la différence de Z, pour la règle de monotonie $\frac{p \rightarrow q}{\Box p \rightarrow \Box q}$, qui caractérise les logiques admettant une sémantique de Scott-Montague. Le miracle plus haut doit être un peu relativisé.

La mauvaise nouvelle n'est pas si mauvaise que ça, mais il faudrait entrer dans des considérations plus techniques pour étayer cette proposition. Ce qui est surprenant ici est que la formule B n'est pas démontrée par Z_1. L'aspect "quantique", dû à B est bien lié alors à une notion de troisième personne du pluriel. La constante *empirique* de Planck, qui définit le niveau où les phénomènes quantiques sont incontournables, définirait le niveau de la duplication où nous survivons en tant que populations de machines.

Les logiques Z_1 permettent de formaliser un certain nombre de questions naturelles sur la *phénoménologie* de la matière isolée ici. Viole-t-elle les inégalités de Bell ? De quelle logique quantique s'agit-il ? S'agit-il de la logique de Birckhoff et von Neumann 1936 ? Définit-elle une machine universelle quantique ? Il s'agit de problèmes ouverts, comme celle de l'axiomatisabilité finie de toutes les logiques Z.

On peut espérer extraire une sémantique algébrique des logiques Z_1* de la forme d'un treillis des sous-espaces d'un espace de Hilbert. Dans ce cas on pourra utiliser des résultats d'unicité de mesure pour extraire une formulation à la Feynman de la mécanique quantique purement déduite du discours de la machine universelle qui s'observe elle-même. On pourra alors *commencer* à chasser le lapin blanc[27] et le cochon volant ...

[27] On trouvera une discussion intéressante sur le net sur différentes stratégies pour chasser le lapin blanc dans le cadre des métaphysiques acceptant l'existence de tous les mondes possibles à l'adresse http ://www.escribe.com/science/theory/.

La preuve ayant été vérifiée indépendemment par de nombreuses personnes, je commence à tenir le résultat du renversement pour acquis. Je pense de même que la stratégie de l'interview de la machine universelle et de son ange gardien est naturelle. C'est la différence entre les logiques des propositions communicables (basée sur G) et celles contenant des propositions vraies incommunicables (basées sur G*) qui permettent de clarifier de nombreux points obscurs en philosophie des sciences et de l'esprit.

Ce dont je suis moins sûr est la pertinence du présent choix des définitions théétètiques de la connaissance et de l'observation. D'autre choix sont possibles. En particulier on peut réappliquer l'idée de Théétète et définir une nuance intensionnelle (modale) très faible de prouvabilité en étudiant la logique des propositions prouvables, consistantes et vraies. Ce qui est étonnant, c'est que la restriction Σ_1 sur S4Grz collapse toutes les modalités : cela ne donne que le calcul propositionnel[28]. Mais avec cette double itération de l'idée de Théétète on obtient à nouveau une logique "quantique" qui prouve la formule B (voir la thèse) et qui permet de capturer des notions de sensations physiques ou de qualia[29] "vraies".

De plus on pourrait s'intéresser à toutes les logiques obtenues avec les Σ_α restrictions, avec α ordinal constructif de Church et Kleene (l'équivalent constructif des ordinaux de Cantor). Il n'en reste pas moins qu'il est étonnant qu'une traduction aussi "brute" de l'argument du renversement isole aussi rapidemment une interprétation arithmétique de la formule B. C'est d'autant plus étonnant que l'idée de Théétète a d'abord conduit à S4Grz dont la sémantique de Kripke est antisymétrique. On aurait pu craindre s'écarter d'une logique de la physique où l'on aurait plutôt besoin d'une sémantique de Kripke symétrique. La symétrie, dont Maria Louisa Dalla Chiara, logicienne quantique de Florence dit qu'elle est bienvenue pour une logique des propositions physiques, préserve l'idéalisme quantique ou computationnaliste du subjectivisme ou du solipsisme.

Notons encore que le fait que la couronne $Z_1^* \setminus Z_1$ est non vide permet d'expliquer pourquoi les sémantiques des logiques quantiques peuvent servir pour axiomatiser la notion de *sensation* physique et les qualia (cf aussi Bell J. L. 1986) car il s'agit d'observables (mesurables) incommunicables (pensez à un plaisir ou à une peine).

---

[28]Ceci est incorrect. Parce que l'on doit tenir compte de l'affaiblissement de la règle de substitution après la restriction Σ_1. Merci à Éric Vandenbussche pour m'avoir signalé cette erreur. Comme $S4Grz_1$ prouve la formule B, et est fermé pour la règle de nécessitation, il pourrait constitué un nouveau pointeur vers une logique quantique arithmétique.

[29]Terme utilisé dans les sciences de l'esprit pour désigner le contenu phénoménologique de la sensation physique.

L'idée de m'inspirer du Théétète de Platon viendra de ma réflexion sur les rêves[30] . Ce qui m'a frappé est la dissymétrie existant entre l'état de rêve et l'état d'éveil : quand on est éveillé on ne peut jamais en être vraiment sûr, par contre, en état de rêve on peut quelques fois s'en apercevoir[31]. Presque tout le travail était terminé en 1986. Je développerai cependant *l'arithmétisation* des définitions de la connaissance de Théétète à l'Université Libre de Bruxelles, à l'IRIDIA plus précisément, grâce à un projet national de recherche, et c'est l'occasion de revenir à l'histoire de la thèse. Jusqu'à présent, la "thèse" n'était qu'un hobby. Je cherchais à répondre à des questions que je me posais. J'avais bien l'idée d'écrire un jour des articles ou un livre, mais je ne croyais plus, depuis 1977 (voir le chapitre 4) à l'idée de faire une thèse *académique*.

[30]Voir le rapport technique IRIDIA 1995. Il contient un chapitre très détaillé sur la nature des rêves, ainsi que des extraits de mes carnets de rêves. Depuis 1976 je note mes rêves nocturnes.

[31]On parle alors de rêve lucide. Le rêve lucide a été mis expérimentalement en évidence par le parapsychologue Hearne, et puis par le mathématicien neurophysiologue LaBerge dans les années 1980. Voir LaBerge 1991.

Chapitre 9

IRIDIA, *mon amour* (1987 → ...)

"Verhofstadt ! Verhofstadt ! ...."

Je me demandais bien ce qui prenait à mon ami le professeur Philippe Smets de prononcer avec enthousiasme le nom d'un ministre libéral flamand bien connu, en Belgique.

- "Verhofstadt" tu es au courant ? Insiste-t-il ? Tu veux toujours faire de la logique modale ?

-Euh ... Oui, mais quel rapport ?

- 'Projet Verhofstadt, 2 ans renouvellables, autant dire 4, pour de la recherche fondamentale en Intelligence Artificielle. Pose-ta candidature tout de suite, j'ai besoin d'un logicien *modale* à l'IRIDIA'. 'Je pourrai faire de la recherche fondamentale ?' demandai-je, ravis. 'Absolument' me répond Philippe ; tu peux même faire une thèse de doctorat.

Nous nous étions rencontré Philippe et moi lors d'une conférence à la société belge de logique et de philosophie des sciences. Philippe était un médecin qui travaillait sur le problème de l'automatisation du diagnostic médical. Il s'était spécialisé en statistiques médicales et il était convaincu de la non pertinence des statistiques et des probabilités dans ce domaine. Il s'intéressait à la théorie des "fonctions de croyance", de Dempster et Shafer, auquel il contribuait. Il était persuadé de la pertinence de la logique pour formaliser des aspects de cette théorie, et il insista pour que je vienne donner un cours de logique modale à l'IRIDIA.

L'IRIDIA, était l'Institut de Recherche Interdisciplinaire pour le Développement de l'Intelligence Artificielle que Philippe venait de fonder à l'ULB. L'institut n'existait que depuis quelques mois.

Au même moment la patrone de l'unité de conformation des macro-

molécules biologiques (UCMB) Shoshana Wodak, me fit savoir que le siège central de Plant Genetic system (PGS) augmente la pression pour avoir rapidemment des résultats tangibles et mon projet à long terme se transforme en un projet à court terme. Michel pense quitter l'UCMB, fonder sa société, ce qu'il fera, et insiste pour que je l'accompagne et que je serve de consultant au moins à mi-temps dans sa société. Il me fait remarquer que je peux être consultant et en même temps bénéficier du projet de recherche fondamentale de Verhofstadt.

Je décide de quitter PGS, d'abandonner *ANIMA* et de refuser l'offre de Michel : je suis scrupuleux, je sens que la recherche fondamentale que je ferais à l'IRIDIA serait incompatible avec la consultance. C'était la première fois dans ma vie que j'allais être financé pour de la recherche fondamentale, je ne voulais pas prendre le risque d'être perturber par des questions trop pratiques qui risquaient de me distraire.

En janvier 1987 j'entre à l'IRIDIA. 'Au sujet de la thèse, Philippe, n'y compte pas trop'. 'Comme tu veux' me dit Philippe. Il était *cool* Philippe, il avait le *tao* du patron efficace. Les chercheurs de l'IRIDIA avaient toute latitude et étaient mûs par leur enthousiasme naturel fortifié par des discussions et des brain-storming assez réguliers. Chacun fixait son horaire comme il lui plaisait, et évidemment tout le monde travaillait entre 10 et 12 heures par jours. la qualité de l'ambiance était reflètée dans la qualité de la production. Par ailleurs l'IRIDIA était indépendant de toutes les facultés, ce qui garantissait la liberté nécessaire aux franchissements interdisciplinaires. Il y avait une salle café et une salle ping-pong ; en vérité l'IRIDIA était un petit paradis pour les chercheurs. Ceux-ci se finançaient eux-mêmes par des projets Européens de type ESPRIT, où des projets nationaux ou internationaux privés.

J'enseignai la logique modale et la sémantique de Kripke. Je devins le "monsieur Kripke" de l'IRIDIA. Je m'en donnais à cœur joie. J'enseignai aussi les rudiments d'informatique théorique et d'intelligence artificielle théorique selon les travaux de Blum, Case et Smith ... C'est lors d'une assez longue discussion, qui va durer un an, que Philippe Smets et moi arriverons à la conclusion que la logique modale KD formalise certains aspects des fonctions de croyance, quelque chose que Philippe développera en détail avec Natasha Aleshina, une mathématicienne Russe qui travaillait à Amsterdam et qui avait trouvé un résultat similaire de façon indépendante (sur base de travaux de Fattorosi et Barnaba qui utilisait KD pour l'approche modale des probabilités). C'est un point assez important qui me motivera ultérieurement pour la version faible de la connaissance "théététique" utilisé dans le chapitre 5 de la thèse de Lille (voir aussi le chapitre précédent).

Avec l'ouverture d'esprit et la bonne humeur à l'IRIDIA je fini par faire un

exposé sur l'utilisation de la logique modale dans la théorie de l'autoréférence gödélienne et j'exposai à la fin les rudiments de ce que j'appelais la "psychologie exacte des machines". L'exposé fut accueilli très chaleureusement. Philippe me demanda si c'était original. Je lui explique alors le long et vivant débat qui opposait les chercheurs dans le domaine des relations entre le théorème d'incomplétude de Gödel, les machines et l'esprit, en remontant à 1921 (Emil Post). Certains, comme Lucas[1], pensent que le théorème de Gödel réfute le Mécanisme. D'autres, comme Webb et moi-même, pensent que le théorème de Gödel serait plutôt une chance pour le Mécanisme : le théorème de Gödel confirme la thèse de Church et celle-ci protège le Mécanisme des nombreux réductionnismes dans lesquels on l'enferme usuellement.

J'ajoutai alors 'Ceci dit, j'ai peut-être un résultat original, *vraiment* original, du genre faux ou révolutionnaire (j'y peux rien). 'Quoi ça ?' Je lui dis : 'une preuve que si nous sommes des machines alors il n'y a pas d'univers, l'apparence de l'univers, et même *des* univers, serait explicable par la géométrie des calculs possibles des machines possibles, vues par ces machines'. Philippe me propose alors d'exposer la "preuve" à l'IRIDIA.

Paul Gochet, professeur de logique à l'université de Liège, assistait depuis un certain temps à mes exposés de logique. Encore une fois il vint m'écouter. Je lui dis de suite qu'il risquait d'être déçu parce que je n'allais pas parler de logique, pas directement en tout cas. Je comptais exposer le paradoxe du graphe filmé et le paradoxe RE[2]. Il s'agit d'une ancienne version de l'argument du déployeur universel. Je n'aurai même pas le temps d'aborder le paradoxe RE. Mon exposé sera suivi d'une discussion qui s'étendra la soirée et une partie de la nuit. Gochet me fit savoir qu'il avait été très interessé et surpris et me proposa d'envoyer un article au congrès de Sciences Cognitives à Toulouse, ce que je fis aussitôt. J'avais une immense estime pour Gochet, encore impressioné par les discussions avec mon ami Dominique sur son "Esquisse d'une théorie nominaliste de la proposition". Mais j'étais étonné, Paul Gochet est une espèce rare de logicien belge, passionné de philosophie analytique, spécialiste de Quine. Les philosophes analytiques, surtout à cette époque, dissolvaient stérilement les questions de philosophie de l'esprit. En fait Paul Gochet avait compris que mon approche conduisait à une reformulation purement mathématique du problème du corps et de l'esprit, et, qu'avec les logiques de l'autoréférence, je construisais un modèle où ces questions avaient un sens mathématique, arithmétique même. Il m'avait écouté et il avait compris que j'avais mis le doigt sur quelque chose[3]. Philippe Smets

[1]On est en 87, Penrose n'a pas encore oublié "The Emperor's New Mind" qui relancera le débat et l'étendra à la communauté des physiciens.

[2]RE est mis pour Récursivement Énumérable, c'est-à-dire générable par ordinateur.

[3]Paul Gochet est le premier à avoir compris que je transforme le problème du corps et

avait compris que ma démarche était sérieuse mais il était extrêmement sceptique aussi bien sur les résultats que la portée pratique des résultats, il estimait cependant qu'il y avait là largement de quoi faire une thèse, et, donc reviendra à la charge sur cette question : "il n'y aura pas trente-six projets Verhofstadt, tu sais, c'est l'occasion ou jamais".

A Toulouse, où mon papier sera accepté, je serai très bien reçu, tout le monde me dira de publier et de faire une thèse, et à Bruxelles idem. Cela devenait fatiguant. A vrai dire cela devenait angoissant. J'irai dire à Smets que je voulais bien faire une thèse à condition que je la dépose et la soutienne à Liège, ou à Louvain, ou à Toulouse : pas à Bruxelles.

'Pourquoi ?' 'Disons qu'à la faculté des Sciences, le département de mathématique n'est pas très ouvert sur Gödel, et le département d'informatique n'est pas très ouvert sur l'intelligence artificielle". Philippe me dit qu'il est au courant : il y avait effectivement, comme on pouvait s'y attendre, une guerre froide entre l'IRIDIA et le département d'informatique. Celui-ci jalousait et critiquait en même temps l'IRIDIA. 'Mais', ajoute-t-il, 'il ne pourront rien contre toi, les jurys de thèses à l'IRIDIA sont étendus avec des experts étrangers, il suffit de t'écouter pour voir que tu argumentes, ils ne se ridiculiseront pas en public : que crains-tu ? Qu'il trouve une faille dans ta preuve ? Ecris d'abord ta thèse, on verra bien ensuite, tu grossis les difficultés'.

Je ne lui ai pas parlé de X, ni du calvaire que j'avais enduré. Il ne comprendrait pas. Cela concernait un travail de fin d'étude, cela datait de presque 20 ans. Je devais grossir les difficultés, peut-être.

Et donc, je commencerai à rédiger cette thèse. Autant j'apprécie argumenter avec un auditoire, autant je déteste écrire. Pour convaincre, j'ai besoin de voir les yeux de ceux à qui je m'adresse. Je suis conscient que mes propositions ont un air trop paradoxal. Oralement, je peux m'interrompre et demander s'il y a une objection. Il y en a toujours. Alors je clarifie, je supprime des ambiguités. On fait savoir alors qu'on a compris et je peux passer à l'étape suivante. Pour mes écrits je doute de la patience du lecteur. S'il a une formation scientifique il ne prendra pas au sérieux les passages qui par la nature de mon travail, ont "un air philosophique[4]". S'il a une formation philosophique, il sautera les passages techniques. A qui je m'adresse ? J'essayerai de satisfaire tout le monde et donc j'écrirai la thèse, de 1989 à 1992 : "Conscience et Mécanisme", 300 pages.

Entre temps le personnel de l'IRIDIA circulait rapidemment. La grande majorité était des italiens à présent. Il y avait aussi des chinois de la république

---

de l'esprit en une recherche d'une justification de l'apparance de la matière, et que cela mettait en doute le statut fondamental des sciences physiques.

[4]La philosophie est rangée d'office dans les disciplines littéraires dans nos contrées francophones.

populaire de Chine, des chinois de Taïwan, des vietnamiens, des français, des anglais, des allemands, et une minorité de Belges. Par une sorte de magie dont Philippe était en grande partie responsable, l'atmosphère était semblable à celle du début sereine et enthousiaste.

De 1992 à 1994 je réécris certains chapitres, j'approfondi considérablement le chapitre sur les rêves et l'analyse Gödélienne du cogito de Descartes. Je veille à ce que les références soient exactes, je décris les tours et détours de mes prédescesseurs dans le labyrinthe des interférences entre Gödel et la philosophie computationnaliste, je rajoute les nombreux programmes LISP qui illustrent toutes les notions techniques de la thèse, comme les "amibes" les programmes miroirs, les "planaires", les démonstrateurs de théorèmes des logiques modales utilisées et le déployeur universel lui-même. La thèse grossit de 300 pages à 750. Je prétexte, en réalité n'importe quoi pour ne pas devoir la déposer. Je commence à ressembler à ces pauvres chercheurs qui semblent être incapables de terminer une thèse. Un beau jour, fin 1994, Philippe *m'annonce* que la thèse est terminée, et que je dois la déposer.

'OK', lui dis-je, 'je vais la déposer à Louvain'.

(Je rêvais de déposer ma thèse à l'Université Catholique de Louvain car cela donnerait un label de sérieux à mon chapitre de théologie, et puis, surtout, je m'étais toujours très bien entendu avec les logiciens, aussi bien mathématiciens que philosophes de Louvain, comme Ladrière qui en particulier était à la fois philosophe et mathématicien, mais aussi Thierry Lucas et Marcel Crabbé. Je me doutais qu'à Louvain, cela ne serait pas une partie facile, en aucune façon gagnée d'avance, mais j'étais sûr que cela engagerait un débat profond et intéressant, et pour moi c'était la seule chose qui compte. La contradiction apparente entre le mécanisme et le catholicisme était due à mon avis à la toujours prévalente conception prégödélienne et réductionniste de la machine[5]).

[5]Pour un exemple extrême voir Jacques Arsac, Les machines à penser, des ordinateurs et des hommes, Seuil, 1987. Bien évidemment le computationnalisme, tel que je le définis est beaucoup plus platonicien qu'aristotélicien. Nous sommes plus proches de Jean Trouillard : L'Un et l'Âme selon Proclos, Éditions Les Belles Lettres, Paris, 1972) que d'André Léonard "Foi et Philosophies, guide pour un discernement chrétien" culture et vérité, Namur 1971, ou "Sciences et théologie, les figures d'un dialogue" de Dominique Lambert au Presses Universitaires de Namur, 1999. J'apprécie l'appel au dialogue entre scientifique et théologien, mais un tel dialogue ne doit pas être utilisé pour minimiser l'importance du Platonisme et exclure d'éventuelles théologies (et théotechnologies) de nature analytique ou déductive. Dans l'annexe sur la thèse de Church je vais jusqu'à suggérer que la thèse de Church réhabilite la philosophie de Pythagore, abstraction faite de ses aspects les plus superstitieux. Pour une introduction moderne à Pythagore, voir Dominic J. O'Meara "Pythagoras Revived", Clarendon Press, Oxford, 1989. Je ne résiste pas non plus à citer

'Ce serait une gifle pour l'ULB', me répondit Philippe. Je lui ai répondu que si je la dépose à l'ULB, ce sera plutôt une gifle pour moi, ou pire : il m'obligeront à la réécrire 10 fois, ils la raboteront jusqu'à ce qu'il n'en reste plus rien, etc.

Philippe me fait savoir alors qu'il en a par dessus la tête de ma "paranoïa". Et, je constate, qu'effectivement j'ai l'apparence du parfait paranoïaque. Comment puis-je lui donner tort ? Suis-je parano ? Intellectuellement je voulais bien le croire, mais dans les tripes, je sentais que j'allais droit à l'abattoir.

le très beau petit livre de O'Meara, traduit en français : Plotin, une introduction aux Ennéades, Cerf, Éditions Universitaires de Fribourg, 1992.

Chapitre 10

Plus noir que vous ne pensez [II] (1995-1998)

Le refus de communication directe est l'arme absolue des pervers.
Marie-France Hirigoyen.

Début 1995, je dépose la thèse à Bruxelles. Je ne voulais pas peiner et énerver davantage Philippe, et après tout c'est grâce à lui et à l'IRIDIA, organisme de l'ULB, que j'avais écrit ce travail. En outre, le président du département de mathématiques, un professeur de mathématiques avec lequel je m'étais fort bien entendu pendant mes études et à qui j'avais offert un exemplaire de la thèse, m'avait presque rassuré. Lui ayant fait part de l'existence d'un problème avec mon mémoire de fin d'étude, il me fit remarquer que cela s'était passé il y avait plus de 20 ans, que le vent avait tourné, que l'intelligence artificielle était prise au sérieux, et que personne au département ne prendrait le risque de se ridiculiser devant des experts étrangers : que je n'avais rien à craindre, qu'il y veillait personnellement, etc.

Pour des raisons de politesse académique il me propose fermement d'offrir un exemplaire de thèse à X, et de lui demander d'être promoteur "officiel" (comme cela se fait), Philippe Smets serait mon promoteur réel "officieux".

Et donc je vais proposer la thèse et la proposition du président à X ; il accepte l'exemplaire sans broncher. 'Je vous tiendrai au courant' finit-il par dire en me montrant la porte. 4 jours plus tard, par e-mail, il dit qu'il refuse d'être le directeur de ma thèse. Tant pis ! Tant mieux ! Ouf !

Une personne bien intentionnée me suggère d'offrir un exemplaire à, un certain Y, disons. Il a une bonne réputation, il donne des cours avec succès

sur l'histoire des mathématiques, et je me souviens aussi avoir fort apprécié son cours de théorie des nombres. Des bons souvenirs.

J'ai une légère hésitation. A des amis, autour d'un verre, je confie que je crois que Y était communiste dans les années 70.

1973-77, ce sont les années de mes études. Je n'avais pas la réputation d'être vraiment *communiste* à cette époque. Pire j'appartenais à cette gauche orwélienne qui n'avait jamais cru au succès de la *révolution*. Je le dois à divers facteurs. Simon Leys est le frère de l'associé de ma grand-mère aux éditions Larcier et on m'a offert tous ses livres si lucides sur la chine[1]. Mon père était militaire et, je dirais, un vrai *juste* à mes yeux, et donc un contrexemple vivant et proche contre le discours antimilitariste primaire. Mais surtout, en 68 j'étais trop petit, et j'attendais l'arrivée du Watson.

Mes amis et moi-même éclatons de rire et chantonnons 'parano ! parano !' 'Ok, ok, les amis, j'irai offrir un exemplaire à Y'. Soupir.

Le hasard intervint et me fit rencontrer Y dans un parking.
- Monsieur Y je voudrais vous offrir un exemplaire de ma thèse...
- Euh ... c'est que ..., non merci, ...
- pourquoi ? m'exclamai-je spontanément.
- c'est que ... X m'a dit que c'était mauvais.
- ..., et vous ne voulez pas vous faire votre propre opinion ? m'exclamai-je tout aussi spontanément !

Il s'en va, l'air vexé. 'Mince alors' pensai-je.

Philippe fut un peu étonné et déçu du comportement de X et de Y, et fut étonné de mon soulagement. J'étais soulagé de voir X s'écarter du chemin.

On décide de suivre la procédure typique dans ce genre de situation qui consiste à proposer une liste de professeurs susceptibles de constituer un jury, avec des experts étrangers (étranger au département de mathématique de la faculté), c'est-à-dire une dizaine de professeurs de l'ULB (physiciens, ingénieurs, biologistes, etc.) et une demi-dizaine de professeurs extérieurs.

Résultat : tollé et hystérie aux département de mathématiques et d'informatique. Je ne comprend rien. De longs mois de suspense. On me dit d'attendre, que cela va se calmer. Un problème avec ma thèse ? 'Rien à voir' m'explique-t-on. Il faut seulement ménager des susceptibilités.

[1] "Les habits neufs du présidant Mao", Ombres chinoises, Images brisées.

Puis le calme, et enfin la bonne nouvelle, on a le champs libre pour constituer le jury.

Puis l'*autre* bonne nouvelle selon Philippe. X et Y ont demandé pour être membre du jury. 'Tu vois qu'ils s'intéressent à ton travail', le vent tourne, ils savent que ton travail aura du succès, ils veulent y participer'. Je n'essaie pas d'expliquer mon sentiment. Je ne sais pas quoi dire.

Puis-je reçois une sorte de notification officielle comme quoi la thèse serait effectivement acceptée, il n'y a plus qu'à fixer la date de la défense privée. Rassuré enfin je commets l'erreur d'envoyer ma thèse aux 50 impatients à qui j'avais promis un jour d'envoyer "ma thèse" et dont j'avais conservé l'adresse. Peut-être me suis-je dit que si je ne les envoyais pas je ne les enverrais jamais ... Envoyer 50 exemplaires de 750 pages c'est un boulot !

J'étais de plus en plus nerveux car je n'avais toujours pas la constitution du jury.

26/9/95. 10h du matin chez moi. Le téléphone sonne. Une secrétaire du département de mathématique de la faculté des Sciences m'apprend la constitution du jury. Le président du département est le président du jury, X est le promoteur puis les membres : Y, Philippe et trois experts supplémentaires (de l'ULB et désignés, *normalement*, par le président). Et personne d'autre ? Ils n'ont pas choisi une seule personne parmi les 15 experts proposés par Philippe. Ca sent le coup fourré.

Mais en même temps, je suis un peu rassuré, parce qu'à une défense privée, entre le président et Philippe je ne vois pas comment X et Y peuvent me démolir (sauf à trouver une "vrai" erreur dans mon travail bien sûr, mais depuis un an il semble que mon travail soit la dernière des préoccupations).

'Qu'est-ce que tu crains maintenant, tu es fâché parce qu'ils vont s'approprier le succès de ton travail ! C'est ça ! Mais c'est ça la gloire mon vieux !' me dit Philippe.

Je ne suis pas à l'aise mais la présence du président du département me rassure.

27/9/95. 14h. Philippe vient de s'en aller aux États-Unis, le téléphone sonne dans mon bureau à l'IRIDIA. Une voix affolée. La même secrétaire du département, j'entend des cris au loin. Elle me dit qu'il y a une erreur dans la constitution du jury : Y est le président du jury, X le promoteur etc. Le président du département avait disparu ! Pas moyen de le joindre au téléphone. A ce moment là je réalise que c'est terminé. Je sens une rage infinie montée en moi et j'envoie un e-mail un peu sec à X du genre "pourquoi ?".

Au retour des Etats-Unis de Philippe, je ne suis pas seul, nous sommes trois à lui expliquer qu'il n'y a aucun doute : la manœuvre est grosse comme un pâté de maison, il n'y aura pas de défense ni privée, ni publique et il devrait refuser d'aller à la réunion concernant la recevabilité de la thèse, car elle y sera jugée irrecevable. Philippe se met en colère : on n'aurait jamais vu ça à l'ULB, d'ailleurs il allait parvenir à les convaincre de la nécessité d'élargir le jury vu le caractère vaste du travail, etc. Philippe croyait encore que je craignais qu'ils volent ma "gloire" ou alors que dans un accès de mauvaise foi ils m'empêchent d'avoir un grade. C'était le pire qu'il puisse imaginer.

'Tu rêves Philippe, cela fait plus de 20 ans qu'ils ne m'ont *jamais* donné la parole, ils ne commenceront pas aujourd'hui'. 'Tu sais très bien qu'ils ridiculisent l'intelligence artificielle, qu'ils méprisent les ingénieurs, les philosophes, et qu'ils ont une haine farouche contre l'IRIDIA', lui explique-t-on.

'Vous exagerez et vous verrez'. Philippe était piégé par ses qualités de confiance et d'optimisme, qualité qui avait fait de l'IRIDIA un si bonne environnement. Il n'imaginait pas qu'on puisse refuser la thèse sans m'écouter au moins une fois, en privé, comme c'est la coutûme à l'ULB. Et donc Philippe y est allé à cette réunion. Et il est revenu blanc comme un linge. Le travail a été souverainement, c'est-à-dire sans appel possible, jugé, effectivement, irrecevable.

Soulagé de ne pas avoir à faire à eux, bien qu'ils perdirent d'un coup, à mes yeux, leur crédibilité scientifique ; leur crédibilité rationaliste même. Soulagé de l'avoir quand-même écrite cette thèse, qui ne demandait qu'à être attaquée. Soulagé enfin qu'ils allaient devoir, à la différence de 1977, signer leur forfait. En effet, la procédure les oblige à rédiger un rapport de non recevabilité. Je ne me faisais pas d'illusion. Le rapport est une formalité, la procédure demande seulement qu'apparaissent le titre de la thèse et les mots "non recevable". Trois semaines plus tard je reçois ce rapport, qui fait moins d'une page, datée et signée, et qui effectivement mentionne le titre de la thèse et le mot "non recevable" et pas grand-chose de plus. Le message est que tout est correct, selon les experts, mais qu'il n'y a pas de résultats originaux (!).

On comprend pourquoi ils devaient éviter tout risque de confrontation orale avec moi.

Ainsi donc, j'ai été jugé sans jamais avoir été entendu (ni lu), et ça à deux reprises par quasiment la même personne, et ses amis, à 20 ans d'intervale.

Pour une question que j'essaye de poser depuis 1963, et un embryon de réponse que j'essaye de partager depuis 1971, cela commencait à bien faire.

Lassitude et épuisement.

Deux professeurs, fort courageusement, me proposeront de présenter ma thèse dans leur service : le professeur Paul Gochet de l'université de Liège et le professeur Jean-Paul Delahaye de l'université de Lille. Ce fut un réconfort intellectuel et moral. Quelques années auparavant, Monsieur Gochet avait envoyé mes publications à Monsieur Delahaye, qui m'avait alors invité à Lille pour discuter de mon approche. J'ai accepté la proposition de direction de Jean-Paul Delahaye. Il m'a suggéré d'être le plus clair possible et m'a poussé à mettre en relief la partie la plus originale, la plus surprenante sans doute : le renversement.

Enthousiaste, Jean-Paul Delahaye publia un article sur mon travail dans la revue "Pour La Science". Avant la défense[2] !

C'est pour des raisons quasi-administrative, 2 ans de DEA en informatique, qu'il faudra attendre 2 ans pour que je défende la thèse le 2 juin 1998 à Lille.

En voyant autant de plaques belges arriver sur le parking à l'université de Lille, un peu inquiet Jean-Paul me demanda s'il s'agissait de mes "opposants". Franchement j'ai souris, il s'agissait bien sûr de mes amis. Jean-Paul ne pouvait pas savoir, je raconte cette histoire pour la première fois. Et tout c'est bien passé. Je me suis bien demandé combien de questions le professeur Paul Gochet me poserait, mais elles étaient toutes intéressantes et il s'est arrêté à six. Je me suis fait un plaisir de lui répondre, ainsi qu'à toutes les questions posées par les membres du jury. Le président du jury qualifia ma prestation de magistral et les membres du jury m'ont chaleureusement félicité, je les en remercie. Je suis heureux qu'ils m'aient écouté.

Bruxelles, le 19 mai 2000.

[2]Le monde des machines, Pour La Science, n 243, Janvier 1998, pp. 100-104.