

# Conscience, Matière et Mécanisme

Bruno Marchal

16 septembre 1998

**Remerciements.** Je remercie Daniel Lehman et Vincent Detours pour leurs lectures et remarques judicieuses.

## 1 Introduction

De nombreux philosophes de l'esprit, ou cognitivistes, pensent que la science contemporaine a résolu le "problème de la matière". Par exemple Dennett affirme dans "Consciousness Explained" que la physique d'aujourd'hui n'est confrontée qu'à des problèmes essentiellement techniques. Outre que cette attitude fait fi des questions fondamentales comme celles que posent déjà les philosophes des sciences (d'où viennent les lois de la physique, pourquoi sont-elles formulables mathématiquement?), cette attitude est tout simplement indéfendable en regard des problèmes conceptuels posés par la recherche d'une interprétation de la mécanique quantique. Aucune unanimité n'existe chez les physiciens quant à la façon dont il faut interpréter "l'amplitude de probabilité" omniprésente dans le formalisme quantique. On peut donc parler d'un "problème *fort* de la matière qui concerne la nature même de la réalité. Dans notre travail nous montrons que, *même* si nous faisons abstraction des données contemporaines de la physique, dès qu'on aborde le problème de la conscience dans le cadre précis de l'hypothèse computationnaliste en sciences cognitives, on est inéluctablement confronté à un problème *fort* de la matière. En fait nous montrons précisément qu'avec le computationnalisme le problème du corps et de l'esprit est transformé en un problème "fort" de la matière dont la solution nécessite une réduction *de principe* de la physique à l'informatique théorique.

Du point de vue ontologique le concept de matière, vue comme une substance indépendante de nous, devient superflu, y compris pour rendre compte de l'existence d'autres *esprits*. Le computationnalisme entraîne ainsi une variété d'immatérialisme qui ne conduit pas au solipsisme.

## 2 L' hypothèse du computationnalisme

L'hypothèse du computationnalisme est la conjonction des trois postulats suivants :

- Le postulat du mécanisme digital et indexical,
- La thèse de Church,
- L'hypothèse d'un minimum de réalisme arithmétique.

Le postulat du mécanisme indexical est l'hypothèse selon laquelle *je suis une machine* dans le sens quasi-opérationnel de *je peux survivre avec un corps artificiel*, cerveau artificiel compris. L'aspect digital est qu'on demande en plus que ce corps soit, toujours *en principe*, digitalisable. Cela signifie qu'il existe un niveau de description de "moi" tel que je survive à une reconstitution de mon corps opérée à ce niveau. Le corps est défini abstraitement par n'importe quelle portion de l'univers telle que je survive à une reconstitution de cette portion. On qualifie cette portion de l'univers de *cerveau généralisé*. Tout ce qui est demandé est qu'un tel niveau de description existe et soit en principe descriptible de manière finie. Le qualificatif "en principe" est important : nous ne demandons pas qu'un humain ou une machine soit à même d'isoler pratiquement et de façon prouvable ce niveau : cela se révélera par ailleurs impossible. L'avantage de cette façon de procéder, avec la notion de cerveau généralisé, est qu'elle nous dispense des débats confrontant l'externalisme et l'internalisme en philosophie de l'esprit : la possibilité que le cosmos dans son entièreté soit le cerveau (c-à-d. que la conscience nécessite l'activité du cosmos entier) n'est pas exclue pour autant que le cosmos entier admette une description digitale. De même nous sommes dispensés de la confrontation entre symboliste et connexionniste. Notons encore que ce postulat présuppose un minimum de psychologie populaire (folk psychology). En effet l'hypothèse présuppose qu'on sache attribuer une conscience à un autre et à soi-même relativement à un changement de situation. Nous illustrerons davantage plus bas cet aspect de l'hypothèse au moyen d'une expérience par la pensée (*gedanken experiment*, voir aussi [11, 4]).

La "thèse de Church" est, dans sa version classique et moderne, l'hypothèse selon laquelle toute fonction calculable est programmable (calculable par un ordinateur). Il y a essentiellement deux types d'argument pour croire que cette hypothèse est plausible. D'une part il y a un argument de type empirique : toutes les tentatives, de sources et d'inspiration pourtant très différentes, conduisent toujours à la même classe de fonctions calculables. Pour donner un exemple la classe des fonctions calculables par un circuit quantique est identique à la classe des fonctions programmables, malgré le parallélisme quasi infini permis par les ordinateurs quantiques. D'autre part

il y a un argument conceptuel donné par le théorème dit de la forme normale de Kleene (voir [10], [23], [12], [14], [15], [16]).

L'hypothèse du réalisme arithmétique est l'hypothèse selon laquelle les vérités de l'arithmétique, comme " $1 + 1 = 2$ " ou l'existence d'une infinité de nombres premiers, mais aussi les vérités des propositions traduisibles dans l'arithmétique, comme la majorité des théorèmes de l'informatique théorique, sont indépendantes de "moi". Je désigne évidemment par "moi" ce que chacun est invité à interpréter par *lui-même*. Par exemple, ceux qui prennent une assurance-vie croient implicitement au réalisme arithmétique.

Ces trois hypothèses sont extrêmement faibles. Elles sont beaucoup plus faibles que des versions du mécanisme critiqués dans la littérature, comme les critiques de Searle par exemple. À ma connaissance, parmi les philosophes rationalistes et monistes (non dualiste) seul le physicien Penrose [19] adopte un point de vue critiquant explicitement l'approche computationnaliste dans les sciences de l'esprit.

### 3 La démonstration

L'objet de notre travail a consisté à illustrer et démontrer que le computationnalisme rend obligatoire la dérivation de l'apparence des sensations physiques et de l'apparence des lois de la physique à partir de l'informatique théorique. Nos travaux n'ont pas consisté à extraire la dérivation proprement dite, bien que nous proposons *aussi* un embryon d'une telle dérivation (voir section 5). Je me propose ici d'esquisser la démonstration. Une expérience par la pensée facilite la compréhension<sup>1</sup>. Malgré son apparence de science-fiction, cette histoire est en principe possible et c'est tout ce dont on aura besoin pour la démonstration.

Monsieur D. souffre d'un cancer généralisé. Il a survécu jusqu'à présent grâce à des greffes d'organes artificiels. Aujourd'hui son cerveau est atteint, et son médecin lui propose de remplacer son cerveau malade par un cerveau digital capable de simuler son cerveau tel qu'il est décrit à un certain niveau. Entre mourrir sûrement dans un délai assez bref ou *peut-être* vivre plus longtemps, Monsieur D. fait son choix, et donne son accord au médecin. On peut voir ici une sorte de pari de Pascal. L'opération se fait sous anesthésie

---

<sup>1</sup>L'expérience par la pensée a l'air de supposer que la description du corps et du cerveau biologique suffit pour survivre à la substitution. Avec le déployeur universel (voir plus bas) on peut se débarrasser de cette apparente supposition supplémentaire. La démonstration fonctionne donc aussi pour la notion plus générale de cerveau généralisé.

et semble réussie. Après un mois de convalescence, Monsieur D. peut rentrer chez lui à la satisfaction de tout le monde.

L'histoire se complique lorsqu'on apprend qu'à l'hôpital le cerveau original n'a pas été détruit, et qu'on découvre en outre qu'il est parfaitement sain (le diagnostic du médecin était erroné). Un certain désordre règne à l'hôpital, et une équipe de chirurgiens d'un département voisin décide de reconstruire un corps artificiel pour Monsieur D. grâce à des données contenue dans l'ordinateur central de leur département. Rappelons-nous que Monsieur D. avait déjà en arrivant à l'hôpital un corps artificiel. Bref, après un mois supplémentaire, Monsieur D. peut à nouveau rentrer chez lui. Et ici il y a un problème, ne fut-ce qu'un problème d'identité civile. Pour fixer les idées, appelons D1, la version de Monsieur D. qui est revenue de l'hôpital après un mois, et D2, la version de Monsieur D. qui est revenue après deux mois. Avec l'hypothèse du computationnalisme, et *en supposant que les substitutions ont été faites au bon niveau*, ni D1 ni D2 ne peuvent, avant de se rencontrer connaître la présence de leurs doubles ou *doppelgängers* respectifs. En particulier, tout deux ont raison d'affirmer qu'ils sont chacun le "Monsieur D. original". Cette histoire illustre la première conséquence du computationnalisme : nous sommes en principe duplicables. Cette propriété a elle-même des conséquences importantes pour la suite de l'argumentation. Ces conséquences sont la non-prouvabilité du computationnalisme, l'émergence d'une forme forte d'indéterminisme, une forme de non-localité, une impossibilité pour tout sujet de distinguer le caractère virtuel ou réel de la réalité, et finalement moyennant à nouveau une hypothèse *extravagante* (dont il faudra se débarrasser) une forme générale de non-physicalité qui nous permettra de conclure. L'hypothèse extravagante permet de modulariser les difficultés.

### 3.1 Non-prouvabilité

Pour résoudre le problème d'identité lié à la duplicabilité, en l'occurrence ici le problème de savoir si D survit en D1 ou en D2, le philosophe Nozick d'Oxford, pour des situations similaires, a proposé une notion de *continuateur le plus proche*. Une telle notion est ambiguë. D1 est temporellement plus proche de D que D2. Mais D2 est physiologiquement plus proche de D puisqu'il a hérité du cerveau original. S'il existait une définition non ambiguë de continuateur le plus proche, alors, avec le computationnalisme ce continuateur serait numériquement duplicable et le problème se poserait à nouveau. À présent on comprend que le computationnalisme n'est pas communicable ou prouvable en se plaçant à la place de D1 ou de D2. En effet de leur point de vue respectif, D1 et D2 savent que l'existence du doppelgänger interdit tout espoir de communiquer le fait qu'ils savent qu'ils ont l'impres-

sion (au moins) d'avoir effectivement survécu à l'opération. En particulier ils savent chacun qu'ils ne peuvent pas convaincre leur doppelgänger que "Monsieur D." a survécu en eux. On sait alors que le médecin aurait menti s'il avait garanti la possibilité de survivre à une telle expérience. Le computationnalisme est donc essentiellement une affaire d'opinion personnelle. Cette non-prouvabilité du mécanisme n'entraîne cependant pas sa non-réfutabilité. Nous montrons que le mécanisme est en principe réfutable et qu'il satisfait ainsi le critère de falsifiabilité de Popper. Le computationnalisme, que j'appelle aussi mécanisme, est donc une hypothèse autant philosophique que scientifique.

### 3.2 Indéterminisme

Le fait qu'il puisse y avoir deux "continueurs le plus proche" numériquement identiques a une autre conséquence remarquable : une forme forte d'indéterminisme phénoménologique. Le computationnalisme entraîne la possible auto-duplicabilité d'un individu. Dans le cas où le candidat à la duplication est conscient que la duplication va être effectuée, il ne peut pas prédire son expérience personnelle future. Toute prédiction serait réfutée pour un de ses futurs répliquants. Donc s'il croit au mécanisme, n'ayant aucun moyen de privilégier un de ses continueurs, il se trouve dans un état d'ignorance complet concernant son avenir proche. Le mécanisme entraîne donc du point de vue de la première personne une forme d'indéterminisme *abrupt*, c-à-d. concernant un résultat de mesure (de localisation par exemple). Cet indéterminisme est distinct de l'indéterminisme dans le long terme que l'on peut extraire de l'incomplétude ou du chaos déterministe. Il est phénoménologiquement semblable à l'indéterminisme quantique qui concerne également directement des résultats *instantané* de mesures.

Ce qui est remarquable, c'est que l'indéterminisme à la première personne résulte complètement du déterminisme à la troisième personne, tout comme dans l'interprétation d'Everett [5] de la mécanique quantique. Cette remarque peut être rendue précise en dupliquant non pas un individu mais une collection d'individus. Au sein des populations concernées, l'indéterminisme est communicable à la troisième personne. Dans l'interprétation d'Everett de la mécanique quantique, on a aussi un phénomène de duplication de population, et l'indéterminisme quantique y est un cas particulier de l'indéterminisme phénoménologique que l'on vient d'extraire du mécanisme.

Le problème du corps et de l'esprit va en grande partie être ramené au problème de la quantification de cette nouvelle forme d'incertitude.

### 3.3 Non-localité

Il n'est pas difficile de se convaincre que le computationnalisme entraîne aussi une forme de non-localité temporelle et spatiale. Le mécanisme rend possible une expérience de téléportation<sup>2</sup>. Dans une expérience de téléportation classique, la description d'un sujet est transcrite au niveau adéquat. Le sujet, préalablement anesthésié, est annihilé. Ensuite "il" est reconstitué quelque part ailleurs, en un ou plusieurs exemplaires. La non-localité temporelle et spatiale correspond au fait que quelle que soit la façon dont on quantifie l'incertitude mécaniste, cette quantification est complètement indépendante du moment et de l'endroit des reconstitutions. Ceci résulte, avec le mécanisme, du fait que, si le niveau de substitution a été bien choisi, l'expérience du sujet est indépendante de toute action "extérieure" qui n'agit pas sur la description physique.

### 3.4 Non-distinguabilité Réel/Virtuel

De la même façon, avec l'hypothèse computationnelle, il n'est pas possible pour un sujet de distinguer à coup sûr si un environnement est virtuel ou réel. Descartes était déjà arrivé à cette conclusion à partir d'une réflexion sur le mécanisme et le rêve. Cela signifie que la quantification de l'incertitude computationnaliste choisie ne peut en aucune façon dépendre de la nature substantielle ou virtuelle de la reconstitution émulée par un ordinateur.

### 3.5 Non-Physicalité

La thèse de Church entraîne l'existence d'un programme capable de générer et d'exécuter tous les programmes possibles. J'appelle un tel programme un déployeur universel. On obtient aisément le résultat de la non-physicalité en adoptant une hypothèse supplémentaire selon laquelle un réel (concret et substantiel) déployeur universel est *intégralement* exécuté dans notre univers. Cette hypothèse est vraisemblablement extravagante à cause du "intégralement". En effet, le déploiement universel nécessite un univers en extension constante avec production ininterrompue de matière. Nous verrons brièvement à la section suivante comment on peut se débarrasser de cette hypothèse extravagante (HE). Comme la quantification de l'incertitude mécaniste ne dépend pas de la nature substantielle ou virtuelle, ni du moment ni du lieu de la reconstitution, dès qu'on admet l'existence d'un déployeur universel concrètement et intégralement exécuté dans notre univers, on est aisément

---

<sup>2</sup>On prendra soin de ne pas confondre la téléportation mécaniste de la téléportation quantique [1].

conduit à une forme radicale d'indéterminisme s'appliquant à toute expérience subjective possible. Considérons par exemple l'expérience simple du lâché d'une craie. Il n'est plus possible d'invoquer les lois de la physique pour inférer la croyance de la proposition "je vais voir la craie tomber" puisqu'il faut a priori tenir compte de toutes les continuations possibles de "mon état computationnel", continuations générées à coup sûr par le déployeur universel. On est dès lors obligé de justifier les lois mêmes de la physique en isolant une façon naturelle de quantifier l'incertitude computationnaliste. Autrement dit on doit extraire les "apparences" des lois de la physique à partir des "croyances les plus probables" des machines. Cela revient encore à isoler une mesure sur les histoires computationnelles contenues dans le déploiement universel, et cela transforme le problème du corps et de l'esprit en une dérivation de la physique à partir de l'informatique théorique.

## 4 La suppression de HE

Pour se débarrasser de l'utilisation de l'hypothèse extravagante, on peut soit utiliser une version conceptuelle assez forte du *rasoir d'Occam*, soit plus proprement, utiliser l'argument dit du graphe filmé [11], soit encore utiliser un argument conceptuellement équivalent dû à Maudlin [17]. Le graphe filmé ou l'argument de Maudlin montre essentiellement le caractère incompatible du computationnalisme avec la thèse de la supervénience physique. Autrement dit il est impossible d'attribuer l'état *privé* de la conscience à l'activité physique (d'un cerveau, d'un univers) dès que l'on retient une forme même faible de computationnalisme. On est dès lors obligé<sup>3</sup> d'associer les états de conscience à des types d'états computationnels dont l'existence, bien qu'abstraite, est indépendante de nous par le réalisme arithmétique. Cela rend superflu l'usage de HE dans la dérivation précédente. Je renvoie aux articles et travaux cités pour plus d'explications.

## 5 L'avis des machines

Si nous voulons résoudre le problème du corps et de l'esprit, nous devons à présent isoler une quantification naturelle de l'indéterminisme du mécanisme à partir de l'informatique théorique. Comment faire ? L'idée que nous avons choisie est de poser la question aux machines elles-mêmes. Ceci est d'autant plus pertinent que nous cherchons à justifier des apparences "physiques"

---

<sup>3</sup>... *si on veut* (à la différence de Maudlin 1989) conserver l'hypothèse du computationnalisme. On peut aussi abandonner à ce stade, comme Maudlin, le computationnalisme !

*du point de vue* des machines. On dérive alors la physique à partir d’une sorte de psychologie, elle-même extraite de l’informatique théorique. Cette “psychologie” est constituée des discours possibles des machines. Toutes les machines n’ont pas nécessairement des choses intéressantes à dire, aussi est-il naturel de se limiter aux discours des machines *auto-référentiellement correctes*. Une machine est auto-référentiellement correcte lorsqu’elle énonce des propositions vraies sur elle-même [13, 21].

Il n’y a bien sûr pas lieu d’aborder dans ce court aperçu les questions techniques que cette approche pose. Les idées de base sont cependant très simples. La communication des machines auto-référentiellement correcte est modélisée à partir de la notion de preuve formelle, c-à-d. la prouvabilité gödélienne. Le sujet et l’objet sont alors définis, par une procédure mathématique au moyen de variantes intensionnelles (modales) de la prouvabilité. Ces variantes intensionnelles apparaissent comme une formulation moderne d’une idée, qu’on peut faire remonter au Théétète de Platon, et qui a une longue histoire dans la philosophie de la connaissance [3, 20]. L’avantage, à mon avis décisif, de l’usage de la prouvabilité Gödélienne va être de fournir, grâce au second théorème d’incomplétude de Gödel et à sa spectaculaire généralisation due à Löb, mais aussi grâce aux travaux de Solovay, une paire de systèmes formels, l’un décrivant les vérités autoréférentielles *communicables*, l’autre décrivant les vérités autoréférentielles *incommunicables* par les machines. Nous rejoignons ainsi [18] lorsqu’il affirme que les théorèmes de Gödel constituent des premiers théorèmes de psychologie exacte. C’est la différence entre la communicabilité et l’incommunicabilité qui va permettre la distinction entre la sensation physique (privée et qualitative) et la mesure physique (communicable et quantitative).

## 5.1 La communication

La communication est modélisée par la prouvabilité formelle. On sait donc reconnaître mécaniquement si une démonstration est valide. En 1976 Solovay a démontré qu’un système de logique modale, appelé parfois G, permet d’axiomatiser adéquatement et complètement, au niveau propositionnel, la part prouvable de la logique de la prouvabilité formelle gödélienne [22], voir aussi [2]. De façon légèrement simplifiée (pour être court) cela signifie que la machine  $M$  prouve  $Bew(\ulcorner A \urcorner)$  si et seulement<sup>4</sup> si G prouve  $\Box A$ .  $M$  est supposée être autoréférentiellement correcte et donc consistante. Comme je l’ai dit plus haut, Solovay a aussi isolé une autre logique, souvent appelée

---

<sup>4</sup> $Bew$  représente le prédicat de prouvabilité de Gödel (*beweisbar*) et  $\ulcorner A \urcorner$  représente un codage de la formule  $A$  dans l’arithmétique.

$G^*$ , qui axiomatise complètement l'entière, au niveau propositionnel, de la logique de la prouvabilité.  $G^*$  prouve tout ce qui est vrai sur la machine  $M$ , pas seulement la part prouvable de la machine (ce que fait  $G$ ). La différence entre  $G$  et  $G^*$  formalise ainsi la part vraie mais non prouvable par la machine de la logique de la prouvabilité. Par exemple le second théorème de Gödel :

$$\neg Bew(\ulcorner \perp \urcorner) \rightarrow \neg Bew(\ulcorner \neg Bew(\ulcorner \perp \urcorner) \urcorner)$$

La non-prouvabilité du faux (c-à-d. la consistance) entraîne la non prouvabilité de la non prouvabilité du faux (c-à-d. la non prouvabilité de la consistance). L'image modale de cette formule,  $\neg \Box \perp \rightarrow \neg \Box \neg \Box \perp$ , est un théorème de  $G$  et de  $G^*$ . Mais la consistance proprement dite,  $\neg \Box \perp$ , est un théorème de  $G^*$ , mais n'est pas un théorème de  $G$ .  $\neg \Box \perp$  est vrai pour la machine, mais n'est pas prouvable par la machine. Avec notre modélisation, il s'agit d'une proposition *incommunicable*.

## 5.2 Le sujet

La logique du sujet, ou de la connaissance du sujet, est obtenue au moyen de l'idée classique [20, 3] consistant à définir la connaissance par l'opinion correcte accompagnée d'une preuve :

$$\Box p \equiv \Box p \ \& \ p$$

La logique modale obtenue en appliquant cette idée à la logique de la prouvabilité est la logique connue dans la littérature sous le nom S4Grz (Grz pour Grzegorzcyk, voir [8]). Du point de vue du sujet la part communicable du savoir est identique à la part vraie :  $S4Grz = S4Grz^*$ . Cette théorie est très utile, elle permet par exemple de réfuter de façon précise l'argumentation anti-mécaniste de Penrose, voir [16].

## 5.3 L'objet

La logique de l'objet est un double raffinement de la logique du sujet. D'une part, l'idée de Théétète est affaiblie : on remplace la vérité par la consistance. C'est une façon de limiter la quantification de l'indéterminisme sur les mondes où on survit. On peut voir là une sorte de Darwinisme arithmétique. On obtient une logique de type probabiliste (voir [6]). D'autre part les propositions sont restreintes à des propositions strictement *vérifiable*, c-à-d. des propositions dont la vérité entraîne la prouvabilité. Cela permet de capturer les états réellement accessibles par le dépoyeur universel. Ce double raffinement produit des logiques modales capturant, au moins partiellement, des

logiques de type quantique. Cela découle d'un résultat dû à Goldblatt sur une analyse modale de la logique quantique [7]. En particulier on obtient une interprétation de la formule modale dite de symétrie  $p \rightarrow \Box \Diamond p$ , où  $\Diamond p$  est une abréviation de  $\neg \Box \neg p$ . Les logiques obtenues sont appelées  $Z_1$  et  $Z_1^*$ . À la différence de S4Grz,  $Z_1$  et  $Z_1^*$  sont différentes, ce qui permet de définir une paire de logiques quantiques, capturant la différence entre l'observation non communicable (le *qualia* de la philosophie de l'esprit) et l'observation communicable (la mesure répétable des physiciens).

## 6 Comparaison avec la Mécanique quantique

Le problème consistant à rendre compte, à partir de la seule équation de Schrödinger (continue et déterministe), de la phénoménologie de la réduction du paquet d'onde (indéterministe) en mécanique quantique est, avec le computationnalisme, un cas particulier du problème du corps et de l'esprit. C'est une conséquence de notre travail. Mieux : au cas où il se confirmerait que la mécanique quantique (plus exactement la linéarité de la mécanique quantique) serait universellement valide, alors le problème du corps et de l'esprit serait résolu par une phénoménologie de la réduction du paquet d'onde, comme celle proposée par Everett. En outre, les expériences par la pensée *et* les sémantiques des logiques de l'observation extraite de la logique de la prouvabilité rencontrent, à nouveau l'interprétation d'Everett, mais aussi les versions plus modernes de cette interprétation comme celle de Isham [9], où les probabilités sont définies en terme de sommes pondérées sur des histoires computationnelles possibles. Et ici encore le computationnalisme rencontre la mécanique quantique (toujours avec les formulations *à-la Everett* c'est-à-dire essentiellement les formulations sans réduction du paquet d'onde) en ce que l'indéterminisme est quantifié aussi bien sur des observables que des inobservables. L'apparition d'une interprétation non triviale de l'axiome de symétrie ( $p \rightarrow \Box \Diamond p$ ) nous laisse espérer comprendre pourquoi *l'ignorance ondule*<sup>5</sup>, et pourquoi, comme la théorie et la pratique nous invite à l'inférer et comme l'expérience nous le confirme, il nous est permis *d'exploiter* cette nature ondulatoire de l'ignorance, ce que l'ordinateur quantique illustre de façon fabuleuse<sup>6</sup>.

---

<sup>5</sup>Par exemple lorsque la position d'une particule n'est PAS observée, l'amplitude de probabilité  $\Psi(x)$  (dont le carré donne la probabilité d'observer la particule avec la position  $x$ , *si on l'observait*) se comporte comme une onde.

<sup>6</sup>On trouvera de nombreuses références théoriques et expérimentales sur le net à l'adresse <http://wwwcas.phys.unm.edu/~mnielsen/info/index.html> (Michael Nielsen).

## 7 Conclusions

Le travail illustre la difficulté qu'il y aurait à essayer d'utiliser aussi bien les phénomènes d'incomplétude de Gödel que l'étrangeté quantique pour réfuter le mécanisme, et donc illustre la solidité de cette hypothèse. Il montre aussi que si on veut être rationaliste et mécaniste on n'a pas d'autre choix que d'être immatérialiste et on n'a pas d'autre option pour résoudre le problème du corps et de l'esprit ou même uniquement pour fonder la physique que de dériver la physique d'une "psychologie exacte", laquelle peut être fondée, avec la thèse de Church, sur l'informatique théorique et ses variantes intensionnelles. Le chemin que nous avons parcouru dans cette direction nous conduit vers le formalisme quantique en commençant par son interprétation "beaucoup d'histoires (computationnelles)" (*many histories*), et vers une logique "quantique" décrivant une relation de proximité entre ces histoires computationnelles.

Pour résoudre le problème du corps et de l'esprit avec l'hypothèse du computationnalisme, il reste à isoler une mesure naturelle définie sur l'ensemble de ces histoires.

## Références

- [1] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolski-Rosen Channels. *Physical Review Letters*, 70, 1993. 1895.
- [2] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.
- [3] M. Burnyeat. Socrate et le jury : de quelques aspects paradoxaux de la distinction platonicienne entre connaissance et opinion vraie. In M. Canto-Sperber, editor, *Les paradoxes de la connaissance, essais sur le Ménon de Platon*, pages 237–251. Editions Odile Jacob, Paris, 1991.
- [4] J. P. Delahaye. Le monde des machines. *Pour la Science*, 243 :100–104, 1998.
- [5] H. Everett III. "Relative state" formulation of quantum mechanics. *Review of Modern Physics*, 9(3) :454–462, 1957.
- [6] M. Fattorosi-Barnaba and G. Amati. Modal Operators with Probabilistic Interpretations i. *Studia Logica*, XLVI(4) :383–393, 1987.
- [7] R. I. Goldblatt. Semantic Analysis of Orthologic. *Journal of Philosophical Logic*, 3 :19–35, 1974.

- [8] A. Grzegorzcyk. Some relational systems and the associated topological spaces. *Fundamenta Mathematicae*, LX :223–231, 1967.
- [9] C. J. Isham. Quantum logic and the histories approach to quantum theory. *J. Math. Phys.*, 35(5) :2157–2185, May 1994.
- [10] S. C. Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam, 1952.
- [11] B. Marchal. Informatique théorique et philosophie de l’esprit. In *Actes du 3ème colloque international de l’ARC*, pages 193–227, Toulouse, 1988.
- [12] B. Marchal. Des fondements théoriques pour l’intelligence artificielle et la philosophie de l’esprit. *Revue Internationale de Philosophie*, 1(172) :104–117, 1990.
- [13] B. Marchal. Mechanism and personal identity. In M. De Glas and D. Gabbay, editors, *Proceedings of WOCFAI 91*, pages 335–345, Paris, 1991. Angkor.
- [14] B. Marchal. Amoeba, planaria, and dreaming machines. In P. Bourguine and F. J. Varela, editors, *Artificial Life, towards a practice of autonomous systems, ECAL 91*, pages 429–440. MIT Press, 1992.
- [15] B. Marchal. Conscience et Mécanisme. Technical Report TR/IRIDIA/95, Brussels University, 1995.
- [16] B. Marchal. *Calculabilité, Physique et Cognition*. PhD thesis, Université des Sciences et Technologies de Lille, 1998.
- [17] T. Maudlin. Computation and Consciousness. *The Journal of Philosophy*, pages 407–432, 1989.
- [18] J. Myhill. Abstract theory of self-reproduction. In M. D. Mesarovic, editor, *Views on general systems theory*, pages 106–118. Wiley, New York, 1964.
- [19] R. Penrose. *Shadows of the Mind*. Oxford University Press, Oxford, 1994.
- [20] Platon. *Théétète ou de la science*, pages 83–192. Œuvre de la pléiade. Editions Gallimard, Paris, 1950.
- [21] R. Smullyan. *Forever Undecided*. Knopf, New York, 1987.
- [22] R. M. Solovay. Provability Interpretation of Modal Logic. *Israel Journal of Mathematics*, 25 :287–304, 1976.
- [23] J. C. Webb. *Mechanism, Mentalism and Metamathematics : An essay on Finitism*. D. Reidel Publishing Company, Dordrecht, Holland, 1980.