

Rapport sur le mémoire de thèse présenté
par M. Bruno Marchal :

"Calculabilité, Physique et Cognition"

Le mémoire proposé par M. Marchal se situe à la croisée de l'Informatique Théorique, des Sciences cognitives façon Philosophie de l'Esprit et de la Logique. Il concerne le (l'ancien) problème du corps et de l'esprit et la thèse (plus récente !) du computationnalisme : le cerveau peut-il être assimilé à un ordinateur digital ? Si l'on répond par l'affirmative, sommes-nous conduits à la thèse matérialiste d'une détermination de la conscience par les lois de la physique ?

A contre-courant d'idées couramment admises, la thèse de M. Marchal est que précisément, l'hypothèse du computationnalisme conduit à rejeter celle du matérialisme. De manière extrêmement audacieuse il entend montrer qu'au contraire, si l'on accepte cette hypothèse, la physique ne peut s'appréhender que relativement à la psychologie - c'est-à-dire au "calcul" - ce qu'il formule de manière provocatrice en affirmant que c'est la physique qui est réductible à la psychologie et non l'inverse. Il n'entend pas démontrer que le computationnalisme est nécessaire, et même il montre qu'aucune preuve transmissible de ce type ne peut exister ; mais en poussant à l'extrême les conséquences de l'hypothèse mécaniste, il est amené à la conclusion ci-dessus.

Je distinguerai deux parties dans son mémoire. La première (chapitres 2, 3 et 4) est consacrée à établir sa thèse ; dans la seconde (chapitre 5) il pose les jalons d'un programme de recherche, en quelque sorte corollaire, visant à étudier par des méthodes logiques "quel type de machines nous sommes".

Tentons de donner une idée de la méthode employée dans sa "démonstration". B. Marchal nous propose une expérience de pensée (en fait une série du même type) reposant sur l'idée de remplacer un cerveau en chair et en os (si l'on ose dire) par un ordinateur digital. Un certain nombre "d'observations" sont faites et de conséquences tirées de cette hypothétique situation. En particulier l'auteur envisage une machine universelle (au sens de la théorie de la calculabilité) appelée "déployeur universel" exécutant simultanément toutes les variantes possibles d'un cerveau dans une situation donnée : une craie tombe et le sujet l'observe. On est conduit à accepter parmi les variantes possibles celle où le sujet ne voit pas la craie tomber ou plutôt où "il voit qu'elle ne tombe pas" - il n'est pas aveugle. Il faut alors rendre compte de ce que cette situation est hautement improbable. L'existence de cette distribution de probabilités n'est pas une propriété de la physique mais des "machines-cerveaux". On en tire la conséquence que l'expérience cognitive (ici perceptive) à la première personne n'est pas réductible à la physique elle-même mais dépend de propriétés des machines-cerveaux.

Le problème du corps et de l'esprit devient alors celui d'expliquer "comment les apparences d'univers et de matière émergent des expériences possibles de la conscience, c'est-à-dire, avec l'hypothèse mécaniste, des histoires computationnelles possibles". Et voilà le renversement annoncé. En fait ce premier argument suppose une machine ("déployeur") universelle

exécutant la totalité des expériences possibles et un second argument, présenté dans le chapitre 5, tente de s'affranchir de cette hypothèse.

Quoi qu'il en soit, l'explicitation des modalités des expériences cognitives doit alors être recherchée non dans la matière mais dans la "logique" de la conscience elle-même. C'est à dire - toujours l'hypothèse computationnelle - dans la logique (sans guillemets) de la calculabilité-prouvabilité. La seconde partie du mémoire (chapitre 5) tente de poser des jalons dans ce sens. L'enjeu est de forger une théorie de la connaissance comme croyance vraie et validement prouvée. Qu'est-ce qu'une machine - dont on suppose une dose raisonnable d'introspection - peut dire de son fonctionnement c'est à dire de ses raisonnements ? que peut-elle en dire de manière transmissible (on veut éviter le solipsisme) ? On arrive ici dans la problématique de l'incomplétude gödelienne et B. Marchal nous conduit dans une investigation des différents systèmes logiques modaux de la prouvabilité à la recherche de formalismes adaptés à son propos, en particulier à une "phénoménologie de la matière" qu'il confronte aux tendances de la physique moderne.

Espérons avoir donné une (pâle) idée de la démarche de M. Marchal. Qu'en penser ? En premier lieu ce qui frappe est l'audace et l'originalité de la thèse, ainsi que la qualité de l'argumentation. Si le problème du corps et de l'esprit est plus que vénérable et si l'hypothèse du computationnalisme (ou mécanisme) a fait couler beaucoup d'encre, notamment dans l'orbe des "sciences cognitives" dans les 10-20 dernières années, il me semble que le "renversement" proposé, appuyé sur une argumentation extrêmement serrée et informée, est de la meilleure eau et d'un apport théorique réel sur des questions difficiles des rapports entre mathématiques et physique, entre logique et théorie de la cognition. J'ai également été impressionné par la connaissance remarquable de l'auteur des logiques de la prouvabilité, du recul dont il fait preuve en la matière, et par sa vision "profonde" de la thèse de Church et de ses implications.

Le rapporteur ne dit pas qu'il est pleinement convaincu : d'une part il faudrait une étude beaucoup plus longue de ce texte dense ; d'autre part la problématique des "expériences de pensées" esquissées plus haut qui sont centrales dans l'argumentation n'est pas "sa tasse de thé". D'autres points de vue - peut-être moins "spéculatifs" - sont possibles sur ces questions. Mais cela n'empêche pas de reconnaître une pensée forte, originale et stimulante. Et le rôle littéralement fondamental, au sens le plus fort, attribué ainsi à l'Informatique Théorique ne peut que réjouir l'informaticien.

Je me déclare donc très favorable à la présentation de ce mémoire comme thèse de doctorat en Informatique à l'Université de Lille.

Caen, le 16 Avril 1996

Patrice ENJALBERT
Professeur en Informatique