

caractériser mon esprit et son fonctionnement, ce ne sont pas les neurones biologiques particuliers qui se trouvent dans ma boîte crânienne, mais l'organisation abstraite de leur réseau et des flux électriques et chimiques qui les parcourent.

Le *fonctionnalisme* en philosophie de l'esprit est considéré comme acceptable par nombre de gens : ce serait revenir au vitalisme des siècles antérieurs que de soutenir que la pensée et l'esprit présents dans nos cerveaux le sont à cause du type particulier de chimie ou à cause de propriétés biologiques uniques que seules possèdent les cellules humaines.

Il est beaucoup plus naturel de croire que l'esprit, la pensée, la conscience sont le résultat abstrait de processus abstraits réalisés d'une certaine façon dans les cerveaux des humains, mais qui pourraient être réalisés matériellement d'autres façons, comme le programme de calcul des nombres premiers peut exister et fonctionner de mille façons différentes dans mille machines différentes.

Lorsque nous avons envisagé le *déploieur universel* et que nous avons dit que chacun de nous était dedans, nous avons, sans le dire, utilisé l'hypothèse *fonctionnaliste*. Maintenant que nous en avons pris conscience, la conclusion de l'expérience de pensée du *déploieur universel* s'exprime ainsi :

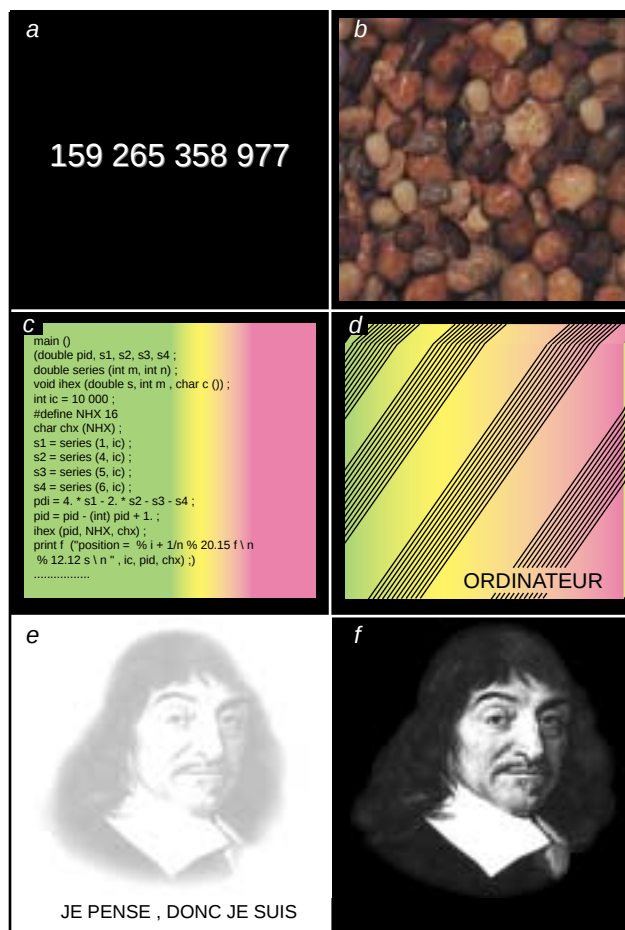
– si nous sommes des *machines numériques* (hypothèse 1) et,

– si le *fonctionnalisme* est exact (hypothèse 2), alors le *déploieur universel* fait fonctionner tous les esprits possibles ayant existé, existants ou qui existeront.

### LA SUPERVÉNIENCE PHYSIQUE

Pour franchir l'étape suivante, nous devons nous interroger sur l'importance des réalisations matérielles nécessaires à l'existence d'un esprit.

Il semble naturel de soutenir que, si aucune réalisation



**3. LA SUPERVÉNIENCE PHYSIQUE.** Un objet mathématique (ainsi l'entier naturel 159 265 358 979) peut être physiquement réalisé : on l'écrit (a) ou on rassemble un nombre de cailloux correspondant (b). Avant que cet entier n'ait été conçu, il existait déjà dans le monde abstrait des objets mathématiques. Il en va de même des programmes informatiques (c) : avant d'avoir été pensé, un programme existe comme un nombre entier. Donc, si je suis un programme (hypothèse mécaniste), mon existence n'exige pas vraiment qu'un monde physique la supporte (e). Cette idée qui dispense d'appuyer la conscience sur la matière est refusée par les philosophes qui soutiennent la thèse de la supervénience physique : pour qu'une conscience existe, il faut qu'un support physique lui soit associé. Cette thèse exige que nous adoptions vis-à-vis des programmes une attitude différente de celle que nous avons vis-à-vis des entiers.

| LE CORPS ET L'ESPRIT                                                                                                                                                                                                         |                                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|
| <b>DUALISME</b><br>(deux types d'entités fondamentales)                                                                                                                                                                      | <b>MONISME</b><br>(un seul type d'entité fondamentale)  |
| Le principe du rasoir d'Occam (il faut être le plus économe possible dans les entités que l'on postule) et la difficulté à concevoir une théorie de l'interaction entre la matière et l'esprit font pencher vers le monisme. |                                                         |
| La double évidence que le monde physique existe et que j'ai un esprit relativement autonome fait pencher vers le dualisme.                                                                                                   |                                                         |
| <b>MONISME MATÉRIALISTE</b><br>(tout résulte du monde matériel)                                                                                                                                                              | <b>MONISME IDÉALISTE</b><br>(tout provient de l'esprit) |
| Le fait que la matière a préexisté à l'esprit, en fait d'apparition récente, fait pencher vers le monisme matérialiste.                                                                                                      |                                                         |
| Les difficultés de la théorie de la supervénience physique font pencher vers le monisme idéaliste.                                                                                                                           |                                                         |

**4. Les oppositions de base entre le corps et l'esprit.**

matérielle ne résulte d'un processus de calcul, non seulement celui-ci ne sert à rien, *stricto sensu*, mais n'existe pas. Cette idée en termes techniques est que l'esprit «supervient à la matière» (qui lui est nécessaire). Autrement dit encore, l'esprit et la conscience naissent de la matière et ne peuvent se passer de ce support.

B. Marchal explore cette hypothèse avec soin et, au terme d'une longue analyse menée avec minutie, il retrouve la conclusion (aussi énoncée par le philosophe T. Maudlin) qu'à partir du moment où l'on accepte le *fonctionnalisme* il faut abandonner la *supervénience physique*. Autrement dit, il n'est pas raisonnable de n'accorder d'existence qu'aux pensées dont une réalisation matérielle existe quelque part dans l'univers physique.

On comprend directement la position de B. Marchal en revenant au cas du programme de calcul des nombres premiers : pour qu'il existe, il n'est pas nécessaire d'admettre que je l'ai vraiment écrit. Il existe en tant que structure mathématique, même si personne n'a jamais vu ce programme, ne l'a jamais écrit ni même jamais envisagé. La conclusion de B. Marchal n'est que la généralisation de cette idée à l'esprit et à la conscience. Une telle conclusion repose sur une option particulière en philosophie des mathématiques appelée le *réalisme* : les objets mathématiques, en particulier les nombres et les programmes (qu'on peut, grâce à des méthodes introduites par Kurt Gödel, ramener à des nombres) existent indépendamment de nous.

Le *réalisme*, quand il ne porte que sur des nombres et des programmes, est difficile à récuser (quoique ce soit l'option retenue par les philosophes *intuitionnistes*), donc la nouvelle hypothèse que propose B. Marchal est plausible. Ainsi, à partir d'hypothèses parfaitement identifiées, B. Marchal défend l'idée qu'un esprit existe en tant qu'être arithmétique abstrait, même si aucune réalisation du programme correspondant n'a été physiquement réalisée ni mise en fonctionnement.

Toutefois, si la conscience que nous avons de nous-même n'a plus besoin de support maté-

1. MÉCANISME NUMÉRIQUE  
(il existe un niveau de détail où nous sommes des machines numériques, c'est-à-dire des programmes)

2. FONCTIONNALISME  
(en changeant des composants d'une machine par d'autres fonctionnellement équivalents, sa nature n'est pas modifiée ; donc mon esprit n'est pas vraiment dépendant de sa réalisation).

3. RÉALISME ARITHMÉTIQUE  
(entiers et programmes existent indépendamment du monde physique)

La solution du problème du corps et de l'esprit est l'idéalisme, qui sera vérifié par notre réussite à déduire la physique de la théorie de la calculabilité.

5. Si l'on admet les trois prémisses de Bruno Marchal, on obtient une nouvelle théorie philosophique et une conception révolutionnaire de la physique. Si l'on pense que la conclusion est invraisemblable,

alors il faut renoncer à l'une des hypothèses au moins. Nombreux sont ceux qui renoncent au mécanisme, mais peut-être que le réalisme arithmétique est l'hypothèse qui doit être abandonnée.

riel, la physique n'est-elle pas une illusion? Dit autrement, la physique pourrait-elle se déduire de l'existence des machines numériques? La physique peut-elle être élaborée purement logiquement à partir de l'étude de la théorie de la calculabilité?

## DUALISME, MATÉRIALISME ET IDÉALISME

B. Marchal défend que cette idée doit être examinée et se fixe le projet de retrouver les lois de la physique par l'étude de la théorie de la calculabilité. Il a déjà obtenu quelques résultats. Avant de les évoquer, je vais situer la conception globale qu'il envisage en la comparant aux théories plus classiques portant sur ce qu'on appelle le problème du corps et de l'esprit.

Très schématiquement, on distingue le *dualisme*, le *matérialisme*, et l'*idéisme*. Le *dualisme* défend que l'esprit et la matière existent et entretiennent des relations l'un avec l'autre. La communication entre les deux mondes est une énigme délicate qui a conduit Descartes (qui soutenait un point de vue dualiste) à proposer qu'elle se faisait grâce à la glande pinéale. Des versions modernes du dualisme ont récemment été défendues par le prix Nobel de médecine 1963 John Eccles, qui propose une théorie fondée sur la possibilité de certains «mécanismes quantiques dans les vésicules du réseau présynaptique des cellules pyramidales du cortex».

Pour des raisons de simplicité et pour éviter le délicat problème de l'interaction entre la matière et l'esprit, le *matérialisme* propose de se débarrasser de l'esprit en décrétant que seule existe vraiment la matière. Dans les versions modernes, le mot *matière* est remplacé par les concepts de particules ou de champs. L'esprit, pour un matérialiste, n'est qu'une illusion ou se ramène à son support physique. Cette conception est celle majoritairement défendue aujourd'hui, car elle semble plus économique en nombre d'entités de base que le dualisme et donc satisfait mieux au principe

du rasoir d'Occam, qui veut qu'en cas de doute, on doive toujours donner la préférence à l'option la plus simple.

Une autre conception satisfait au principe du rasoir d'Occam l'*idéisme* : seul l'esprit existe, la matière n'étant qu'une illusion. À partir du mécanisme, B. Marchal tombe sur un *idéisme*. Son *idéisme* est toutefois très spécial, puisqu'il admet comme élément de base, non pas une substance spirituelle vague et toute-puissante, mais l'univers des *machines numériques abstraites*, qui ne sont qu'un type d'objets mathématiques. Pour cet idéisme, ce qui existe avant tout, c'est une catégorie complexe d'objets mathématiques, et tout se construit à partir de là.

Bien évidemment, la solution de B. Marchal est inattendue, choquante, et ceux qui ont fait l'effort de l'écouter ou de le lire ont rarement réussi à le suivre, tant les deux conceptions dominantes d'aujourd'hui que sont le *dualisme* et le *matérialisme* occupent le terrain et masquent les autres alternatives.

Heureusement sa conception est susceptible d'être testée ou attestée (je n'ose dire prouvée!) : si l'on réussit à déduire la physique de la théorie des machines numériques, la conception de B. Marchal devra certainement être prise au sérieux, quel que soit l'étonnement qu'elle provoque au premier abord. Un premier pas a peut-être été franchi dans cette direction. Pour B. Marchal, la théorie de la calculabilité doit servir de base à une reconstruction de la physique. C'est une théorie non triviale (elle se construit sur l'arithmétique, qui est loin de l'être), dont certaines parties développées pour approfondir la compréhension des théorèmes d'incomplétude de Gödel explicitent ce qu'une machine peut connaître d'elle-même sans se tromper, c'est-à-dire en restant non contradictoire. Cette *logique de la prouvabilité* développée pour comprendre mieux le sens des résultats de Gödel peut être vue comme une *logique des êtres conscients*. B. Marchal en a élaboré une version nouvelle qui semble s'adapter au point de vue qu'il

défend et qui constitue l'apport technique et mathématique de son travail, indépendamment de son interprétation philosophique.

Un point remarquable est que la logique qu'il a développée ressemble à une autre logique qui avait été proposée par G. Birkhoff et J. von Neumann pour rendre compte des observables de la mécanique quantique.

Cette coïncidence doit-elle être interprétée comme le début de la réalisation du programme de reconstruction de la physique par la théorie de la calculabilité? Il est trop tôt pour le dire. Cependant, si c'est le cas, alors cela signifie que la conception philosophique générale qu'envisage B. Marchal doit être examinée avec attention. En explorant avec détermination les conséquences de l'hypothèse du *mécanisme numérique*, B. Marchal a ouvert un champ de recherche nouveau qui, si l'on réussit à y progresser, remettra en cause nos conceptions philosophiques du corps et de l'esprit.

Jean-Paul DELAHAYE est directeur adjoint du Laboratoire d'informatique fondamentale de Lille du CNRS.

e-mail : delahaye@lilfl.fr

B. MARCHAL, *Informatique théorique et philosophie de l'esprit*, Actes du Colloque de l'Association sur la recherche en cognisciences, Toulouse, 1988.

B. MARCHAL, *Mechanism and Personal Identity*, in *Proceedings of WOCFAI, World Conference on the Fundamental of Artificial Intelligence*, sous la direction de M. De Glas et D. Gabbay, Paris, 1991.

B. MARCHAL, *Des fondements théoriques pour l'intelligence artificielle et la philosophie de l'esprit*, *Revue internationale de philosophie*, 172, 1990.

B. MARCHAL, *Calcul, physique et cognition*, Rapport de recherche, LIFL, 1997.

T. MAUDLIN, *Computation and Consciousness*, in *The Journal of Philosophy*, pp. 407-432, 1989.