

## 3.2. Le paradoxe du graphe filmé

### 3.2.1 Introduction

Une formulation détaillée du problème du corps et de l'esprit, ainsi qu'un embryon de reformulation mécaniste et mathématique du problème sont postposés à la section 3.3. Ici je vais plutôt procéder *comme si* le mécanisme présupposait le *matérialisme*, ce qui est motivé par la section précédente sur le rêve. Le mécaniste matérialiste, *non éliminativiste*<sup>1</sup> doit faire reposer l'expérience de la conscience sur l'activité matérielle ou physique du cerveau, du corps ou de l'univers ou identifier cette expérience avec tel ou tel état dynamique. Autrement dit, il doit accepter un principe de supervénience physique ou/et un principe d'identité psycho-neurale, (ou psycho-matière).

Je vais montrer que de tels principes sont incompatibles avec l'hypothèse mécaniste digitale indexicale (MDI), et dans 3.3 je vais montrer comment le problème du corps et de l'esprit se transforme nécessairement en une recherche d'une phénoménologie de la matière avec une ontologie immatérialiste.

Le principe du parallélisme psycho-physiologique ou psycho-physique (PP) postule que la conscience (ou l'esprit) supervène sur (est véhiculé par) l'activité physique du cerveau (du corps, ou de la matière). La section précédente confirme, *en apparence*, ce point de vue y compris en ce qui concerne l'expérience du rêve, en dépit de l'argumentation de Malcolm

---

<sup>1</sup> Pour qui, en particulier, le terme "conscience" admet une référence réelle.

Le rêve est considéré comme une expérience subjective supervenant sur l'activité du cerveau pendant la phase de sommeil paradoxal<sup>2</sup>. C'est un cas particulier de PP compatible avec la conception cartésienne ou cailloisienne des rêves.

L'intérêt du rêve est que le cerveau travaille en se fermant aux mondes qui l'entourent en inhibant les entrées et les sorties (voir 3.1). L'exposé du paradoxe du graphe booléen filmé est plus facile si on le fait reposer sur l'expérience du rêve. Pour ceux qui insistent sur le rêve comme processus chimique ou quantique ou/et s'ils pensent qu'une ouverture sur l'extérieur, peut-être infinitésimale, éventuellement de nature télépathique, soit possible, devront modifier l'expérience par la pensée afin de décrire un graphe émulant le niveau qu'ils estiment adéquat. Le raisonnement marche quel que soit le niveau de fonctionnalisme choisi, et l'existence de celui-ci est donnée par l'hypothèse mécaniste indexicale.

### 3.2.2 Thèses de l'identité

La théorie la plus répandue chez les matérialistes qui se sont intéressés au problème du corps et de l'esprit et qui admettent un parallélisme psychophysique est la théorie selon laquelle on peut identifier les états mentaux avec des états neurophysiologiques correspondants.

L'identification peut se faire au niveau des *tokens* ; par exemple : la douleur que ressent monsieur X à telle date et lieu précis est identifiée à l'état physiologique du cerveau de monsieur X au même date et lieu (hypothèse forte de l'identité).

L'identification peut se faire au niveau des *types*. Un certain type d'état mental est identifié à un certain type d'état fonctionnel ou *dispositionnel* du cerveau et/ou de la matière. On peut alors estimer que le type d'état du cerveau peut être réalisé de manières diverses. Si, avec Putnam 1960, on estime que le type d'état mental peut être attaché avec des réalisations (matérielles et donc notamment en termes de token) du type en question, on est en présence de la position fonctionnaliste traditionnelle (voir 1.1). Les réalisations sont censées rendre compte alors des manifestations (du corps) correspondant à l'état mental considéré.

Si le type est réalisé par une machine digitale, on est en présence du mécanisme digital (et indexical). Dans ce cas le nombre de token est fini et les descriptions locales des réalisations instantanées sont finiment codables.

---

<sup>2</sup> Cela ne signifie pas qu'il y ait absence d'activité mentale ou même de rêve pendant les autres phases. Il semble que lors du sommeil lent une conscience plus verbale qu'onirique, vaguement répétitive, puisse apparaître surtout si l'on se couche avec quelque soucis.

En résumé je distingue deux thèses de l'identité :

La thèse forte selon laquelle (la douleur de X) à l'instant t est identifiée à (l'état physique de X) à l'instant t.

La thèse faible selon laquelle (la douleur de X à l'instant t) est identifiée à (un état computationnel relatif (une collection structurée d'états)).

Ce dernier est a priori atemporel. L'instant de la douleur (par exemple) est inclus dans le quale de la douleur. La thèse forte nécessite au contraire une conception suffisamment physicaliste pour définir une notion de temps réel ou de temps externe.

MEC-DIG n'utilise a priori qu'un principe d'identité faible avec un temps interne (comme celui attaché à la théorie S4Grz dérivé de l'auto-référence), mais si l'on admet un principe de supervénience physique, on a une identité forte et on doit résoudre le problème de la relation entre le temps interne et le temps externe.

Le rôle des token se laisse capturer par diverses sortes de démons. C'est ce qui va permettre la dérivation de la situation paradoxale.

Le paradoxe du graphe filmé montrera que les lois computationnelles de l'esprit sont incompatibles avec la supervénience physique. Pour préciser ces termes je rappelle le principe du fonctionnalisme, et j'exhibe quelques démons.

### 3.2.3 Le principe du fonctionnalisme (rappel)

Je rappelle le principe du fonctionnalisme (voir 1.1)

*Si on isole dans un système A un sous-système B et que l'on remplace B dans A par un sous-système B' **fonctionnellement équivalent** à B relativement à un niveau n, alors on obtient un système A' **fonctionnellement équivalent** à A relativement au même niveau n.*

où B' est éventuellement *un démon*, c-à-d n'est pas nécessairement une machine<sup>3</sup>. Il peut s'agir d'un dieu, d'un malin génie cartésien, d'un concours de circonstances fortuites, d'un oracle, etc.

Pour aborder le problème posé par la supervénience et l'identité, deux sortes de démons seront distingués.

---

<sup>3</sup> Nous avons aussi vu que les processus parallèles pouvaient être émulés par des processus séquentiels, quitte (au cas où seul un mécanisme de bas niveau est correct, à émuler tout l'univers pour capturer le temps réel relatif (dans ce cas-là MEC-DIG-DEI doit être équivalent à PAN-MEC-DIG). Avec un mécanisme de haut niveau, où le soi est purement représentationnel, comme l'onirophysiologie tend à le confirmer, l'émulation du comportement physico-chimique du cerveau devrait suffire pour que la conscience (identique à celle du système non émulé) puisse supervenir.

- 1) le démon contrefactuellement correcte (démon CC), et
- 2) le démon accidentellement correcte (démon AC).

### 3.2.4 Démons contrefactuellement corrects.

#### a) le démon mécaniste

Lorsque l'on remplace une partie par une machine équivalente, c'est-à-dire dans les cas où selon MDI on survit, la nouvelle partie joue le rôle de démon mécaniste. Les prothèses et les greffes artificielles sont des exemples de démons mécanistes. Etant une machine fonctionnellement équivalente (au niveau adéquat), le démon mécaniste est, par construction, contrefactuellement correct : il joue son rôle quelles que soient les exécutions, c-à-d quels que soient les contextes.

#### b) le démon non-mécaniste.

Toute influence extérieure qui ne soit pas mécaniste, mais qui pour quelques raisons *magiques* peut être considérée comme contrefactuellement correcte. Je n'utiliserai aucun démon non-mécaniste de ce genre, en effet MDI entraîne que tout démon CC, dont la présence est nécessaire pour une activité de l'esprit, est une machine.

### 3.2.5 Démons accidentellement corrects.

#### a) pièces relativement défectueuses, ou absence de pièces.

Supposez qu'un automobiliste fasse remplacer par un garagiste une pièce du moteur de son automobile. Par malheur (?) le garagiste présente une pièce qui est encore (partiellement) défectueuse. En effet elle tombe en panne si vous dépassez le 124 km/h. Par chance (?), sur le chemin du retour l'automobiliste ne dépasse jamais le 124 km/h, et il s'estime satisfait de la substitution. Le démon (la pièce substituée) joue son rôle pour un ensemble restreint d'exécutions (celle où l'automobiliste ne dépasse pas le 124 km/h).

Sur ces exécutions la nouvelle pièce est fonctionnellement équivalente à la pièce originale non défailante). Cependant pour chaque exécution, il est contingent (donc accidentel) que l'automobiliste ne dépasse pas le 124 km/h. En ce sens, pour chaque "exécution" où la vitesse est plus petite que 124 km/h le "démon" (qui correspond à la pièce défectueuse) est **accidentellement** correct. Autre exemple : un ordinateur, avec un registre défailant, qui exécute un programme p, est **accidentellement équivalent** au même ordinateur muni du même programme p, mais n'ayant aucun registre défailant, **pour les exécutions particulières** qui n'utilisent pas ce registre.

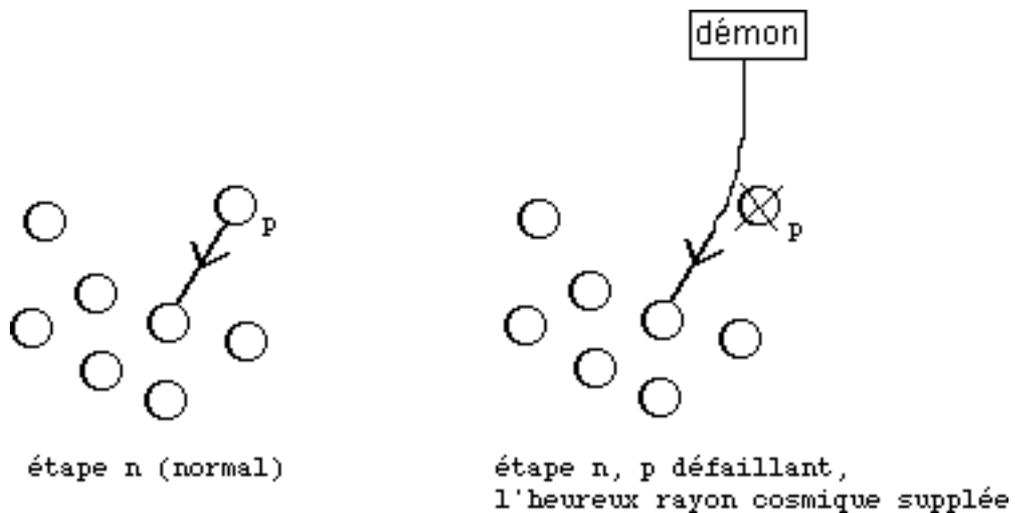
Plus généralement, si pour un ensemble E d'exécutions, une machine n'utilise pas une de ses parties P, celle-ci, inactive lors des exécutions appartenant à E, peut être supprimée (pour ces exécutions). Le démon ici est constitué par l'absence d'une pièce. Il est accidentellement correcte pour certaines exécutions, dans la mesure où la pièce défectueuse n'est pas utilisée pour ces exécutions.

Je parlerai dans ce cas de *démon AC-absence-de-pièce*. Ils seront utiles pour l'analyse de Maudlin 1989.

b) les heureux rayons cosmiques

Voici une autre sorte de démon accidentellement correcte qui va être utilisé dans la suite.

Supposons qu'un neurone, ou une partie d'un ordinateur, doit, lors de l'étape *n* d'une exécution particulière, envoyer un message (une impulsion) à un certain autre neurone (disons). Voir figure suivante à gauche :



Supposons à présent que ce neurone p est juste défaillant à l'étape *n* de cette exécution (après un court instant il est réparé). Il n'a pas envoyé l'impulsion qu'il aurait dû envoyer à cette étape. Supposons à présent, que lors de cette exécution particulière, un accidentel (heureux) rayon cosmique, purement par chance, vient exciter, à l'étape *n*, le neurone qui devait recevoir l'impulsion que p aurait dû envoyer (voir figure plus haut à droite). Dans ce cas, il n'y a pas de différence dans le comportement du système homme ou machine pour l'exécution courante.

*exemple et exercice*: imaginer une conférence d'un professeur au cerveau défaillant, et le faisceau de 100.000.000.000 heureux rayons cosmiques qui suppléent lors de la conférence.

Je parlerai dans ce cas de *démon-AC-heureux-rayon*.

**En résumé**, au niveau adéquat pour MEC-IND, les substitutions sont contrefactuellement correctes (CC) si elles fonctionnent pour toute exécutions possibles, et elles sont accidentellement correctes (pour une exécution ou pour un ensemble restreint d'exécutions) si elles fonctionnent exclusivement pour cette exécution où cet ensemble d'exécutions.

Distinguons dès lors MEC-(DIG-IND) = MEC-(DIG-IND)-CC de MEC-(DIG-IND)-AC, ou plus simplement :

1) **MDI-CC (=MDI)** : je survis (localement) uniquement si les substitutions sont contrefactuellement correctes.

2) **MDI-AC** : je survis (localement) aux substitutions accidentellement correctes.

et on peut distinguer naturellement deux versions de MDI-AC:

*MDI-AC-absence-de-pièce*

*MDI-AC-heureux-rayon.*

*Remarque* On pourrait penser que MDI-AC a été utilisé dans les duplications, car la reconstitution instantanée, par exemple du vieillard dupliqué, peut être considérée localement comme l'oeuvre d'un heureux démon (cf aussi l'accident lointain), mais d'autre part le reconstitué a été supposé, dans toutes les situations décrites jusqu'à présent, être une substitution contrefactuellement équivalente (MDI-CC) vu le caractère 100% fiable des dupicateurs utilisés et l'adéquation du niveau choisi.

### **3.2.6 Les principes de supervénience**

#### *a) la supervénience physique*

**SUP-PHYS** : La conscience supervène sur l'activité physique de la machine.

Je rappelle que *l'activité de B véhicule l'expérience A*, est synonyme, en terme de supervénience, à *l'expérience A supervène sur l'activité de B*.

La motivation principale de la thèse de la supervénience physique SUP-PHYS est la croyance dans le matérialisme. Il y a de la matière quelque part qui interagit causalement, et la conscience serait un phénomène réductible à une catégorie d'activité de cette matière. Matière est pris au sens large de structure spatiotemporelle dynamique. La conscience supervène alors sur une région de l'espace-temps ou plus simplement sur une activité physique

durant un certain intervalle de temps. Une thèse forte de l'identité est nécessaire.

En terme de supervénience physique MDI-AC peut s'écrire :

*Si, lors d'une exécution (et pas conditionnellement pour toute exécution), la substitution d'une partie A par une autre B n'est à l'origine d'aucun changement dans le comportement des entités au voisinage de A, alors si l' exécution sans substitution véhicule une expérience privée l' exécution avec substitution véhicule la même expérience privée.*

Les substitutions étant effectuée au niveau mécaniquement adéquat. Maudlin (1989) démontre (je rappelle que MDI = MDI-CC):

MDI + SUP-PHYS => MDI-AC-*absence de pièce* => une absurdité (en l'occurrence *la conscience supervène sur presque rien du tout*)

#### b) la supervénience computationnelle

**SUP-COMP** : La conscience supervène sur les états computationnels d'une machine.

SUP-PHYS, est cohérent avec le principe d'identité fort (en terme de token).

SUP-COMP, où la conscience supervène sur l'activité computationnelle du *cerveau* (mécanisme de haut niveau), est cohérente avec le principe de représentation de Fodor, avec les théories dispositionnelles des croyances, et avec MDI-CC. Il nécessite une thèse d'identité faible (en terme de type).

Pendant MDI-CC, n'oblige pas de choisir le niveau et est compatible avec une théorie (computationnelle) de l'esprit où la conscience serait véhiculée par l'ensemble complet des états computationnels d'une solution calculable (au sens de Turing) d'une équation de l'univers (comme l'équation de De Witt-Wheeler, il s'agit donc d'un mécanisme de bas niveau).

J'ai démontré avec le paradoxe du graphe filmé, Marchal (1988) :

MDI + SUP-PHYS => MDI-AC-*heureux-démon* => une absurdité (en l'occurrence *la conscience supervène sur rien du tout*)

Notons que MDI-AC => MDI-CC trivialement puisque les substitutions contrefactuellement correctes sont toujours accidentellement correctes.

La section suivante (3.3) abordera les conséquences de SUP-COMP. Je montrerai que le problème du corps et de l'esprit revient à exhiber et justifier, avec SUP-COMP et la thèse de Church, une structure non triviale d'états computationnels immédiatement accessibles, rendant compte (avec SUP-COMP) des probabilités mises en évidence en 1.3.

PGF, en montrant l'incompatibilité entre MDI et SUP-PHYS (et donc de toutes thèses d'identité forte) exige une telle justification.

La preuve de Maudlin a deux avantages sur ma "preuve" de 1988.

1) elle ne donne pas la fausse impression que le paradoxe puisse être pris pour une sorite.

2) elle gère mieux la contrefactualité.

Je vais exposer Marchal 1988, Maudlin 1989, et les comparer. A cette fin je présenterai une version de la preuve de Maudlin en la retraduisant en terme de graphe filmé.

Maudlin (1989) défend le matérialisme, et donc SUP-PHYS, et estime réfuter MDI. Je (1988, et ici), cherchant les conséquences de MDI, suis obligé d'abandonner SUP-PHYS. Dans 3.3, c'est la doctrine du *matérialisme* dont je suggère l'abandon.

Maudlin et moi avons donc montré (indépendamment<sup>4</sup>) que SUP-PHYS et MDI sont incompatibles :

$$\neg \text{MDI} \vee \neg \text{SUP-PHYS}.$$

### 3.2.7 Le paradoxe du graphe filmé

#### 1°) présentation de Macbeth

##### a) le matériel

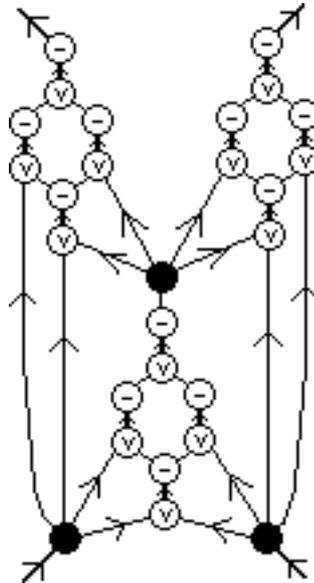
Je présente d'abord une machine universelle particulière. Il s'agit d'une modification du graphe booléen présenté en 1.1. J'ai besoin, pour l'argumentation, d'une machine universelle plane. Je pourrais prendre le jeu de la vie de Conway ou un quelconque autre automate cellulaire plan universel, mais je vais simplement considérer le graphe booléen modifié de telle façon qu'il soit complètement plat.

La seule difficulté revient à faire croiser deux fils dans le plan. Dewdney et ses lecteurs du Scientific American ont montré qu'il était possible de

---

<sup>4</sup> Voir aussi Lycan 1987. On trouvera aussi dans ce livre un argument modal de Kripke contre la thèse forte de l'identité, Kripke 1972. Voir encore MacKay 1960.

réaliser les croisements en utilisant le "nor", c'est-à-dire le "v" suivi du "¬".  
Voici une solution proposée :



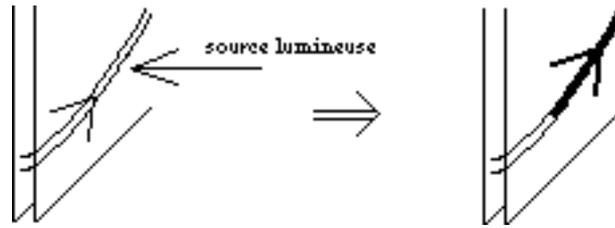
On peut vérifier à *la main* le fonctionnement de ce circuit. La transformation qu'on applique sur le graphe booléen consiste donc à remplacer tous les croisements de fils par ce croisement plat. On ralentit de même l'horloge du graphe pour que le fonctionnement du graphe planaire ne soit pas perturbé par le ralentissement opéré dans ces nouveaux croisements plats.

On voit que les paradoxes qui seront tirés de ce graphe booléen ne seront pas levés par une évocation magique du parallélisme comme on trouve dans quelques approches de la conscience (Johnson-Laird 1987).

En effet la "projection" d'un réseau tridimensionnel en "machine de Dewdney" conserve le traitement parallèle de l'information. Inversement nous verrons que PGF et PDU, généralisant les paradoxes de la duplication de soi, vont permettre de justifier un parallélisme de façon purement interne (logique et computationnelle) corroborant, dès lors sans magie *artificielle*, l'intuition de Johnson-Laird.

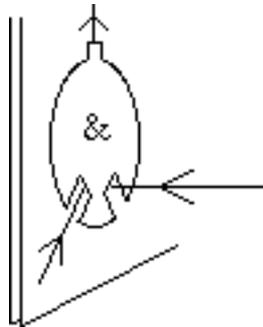
Les modules sont, par définition, constitués des fils (les arcs) du graphe, et des connecteurs.

Les fils du graphe sont des fibres optiques directionnelles et capables d'être excités par une source extérieure perpendiculaire :



Pour le bus on utilise des fibres bidirectionnelles.

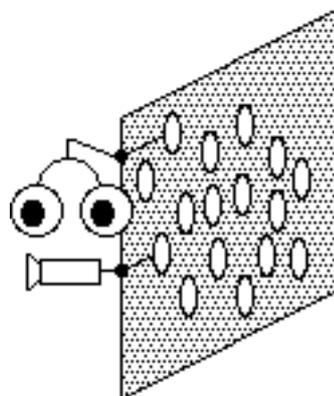
De même les modules "&", "v", "¬" sont supposés être sensibles à la lumière lorsque la direction de la source est approximativement perpendiculaire au graphe. Par exemple voilà le "&". Dans le dessin suivant, il est activé par une fibre optique interne et par une source lumineuse externe :



Il en est de même pour le *ou* (v) et le *non* (¬).

Le graphe est plongé dans de la fumée semi-transparente serré entre deux vitres. De cette façon il nous est possible à la fois de contempler de l'extérieur l'activité du graphe et d'éventuellement le perturber.

Macbeth est le nom donné à ce dispositif. Dans le dessin suivant Macbeth est représenté avec des sorties motrices oculaires et vocales :



## b) le logiciel

Le graphe booléen est initialisé de telle façon qu'il soit fonctionnellement équivalent à l'état instantané (physique, chimique, au besoin quantique) d'un cerveau (d'un corps, etc.), d'une personne que j'appelle Hamlet<sup>5</sup>, en début de phase de sommeil paradoxal, au niveau adéquat (existant par MDI) muni d'organe de sortie, par exemple des cordes vocales actionnées par des impulsions capables d'être transmises par des fibres optiques, ainsi que des globes oculaires actionnés par les fibres optiques.

Par MEC-IND, Hamlet rêve. Si on parvient à capturer l'état du graphe à la fin du rêve et à l'implémenter dans le cerveau original de Hamlet, par MEC-BEH il sera à même de raconter le rêve, et par MEC-DIG-FORT ou MEC-DIG-IND (MDI), son souvenir sera aussi pertinent qu'après un éveil normal. Et cela quel que soit le temps réel mis par le graphe à procéder l'activité du rêve. Supposons en effet que la vitesse du traitement de l'information de Macbeth soit égale à 1/1000 de la vitesse du cerveau original de Hamlet, supposons encore que le rêve de ce dernier soit lucide et qu'on lui ait préalablement demandé de compter jusque 10 en indiquant par des mouvements oculaires le début et la fin de son comptage. Dans ce cas, pendant l'activité de Macbeth, 13000 secondes seront mesurées entre les deux mouvements des globes oculaires de Macbeth (en supposant pour le besoin de l'argumentation une précision accrue dans les résultats de LaBerge voir 3.1). Par le principe de détermination, et puisque durant le sommeil paradoxal le système sensoriel est inhibé (le système est fermé), chaque ré-initialisation de Macbeth sera à l'origine d'un rêve identique, avec déplacements identiques des globes oculaires. Cependant, toutes les fois qu'on interroge Hamlet au sujet de l'originalité de son rêve (soit directement si on a connecté des entrées sensoriels à Macbeth, soit par implémentation de l'état final de Macbeth dans le cerveau original de Hamlet) il répondra affirmativement qu'il a commis ce rêve pour la première fois.

Avec le principe de supervénience physique, le rêve et la conscience de Hamlet superviennent sur l'activité physique de Macbeth.

Insistons sur le fait que le caractère irréaliste de ce dispositif et de sa programmation ne doit pas être tenu responsable des éventuelles difficultés conceptuelles que nous allons rencontrer. Il suffit que MEC-IND soit correct pour qu'une telle expérience puisse *en principe* être envisagée. L'expérience présentée ici est (encore) une expérience par la pensée. De même j'utilise le

---

<sup>5</sup> Macbeth joue ainsi le rôle d'un cerveau artificiel pour Hamlet. Confondre Hamlet et Macbeth revient à confondre deux niveaux, comme Searle 1980, 1984. Voir Hofstadter & Dennett 1981, Maudlin 1989 pour une analyse de cette confusion. (Re)lire 1.1.

rêve lucide pour rendre l'argumentation plus simple et plus saillante. Un simulacron aurait pu être utilisé. On imaginera de nombreuses variantes<sup>6</sup>.

### 2°) première expérience

La première expérience consiste à filmer Macbeth lorsqu'il exécute le rêve. On utilise une caméra capable de restituer des images très précises dans lesquelles on distingue les composants élémentaires (fils et connecteurs) du graphe booléen. Par le *film-de-Macbeth* j'entends la projection sur écran du graphe filmé.

Voici la première question posée :

La conscience supervène-t-elle sur le *film-de-Macbeth* ?

Il semble que répondre oui à cette question revienne à commettre l'erreur la plus grave qu'un philosophe puisse faire : confondre la réalité et un dessin animé (Marchal 1988). Dans un dessin animé un mouvement (apparent) de projectile envoyé sur une vitre n'est pas la cause de l'éclatement (apparent) de la vitre. Le dessin animé est a-causal, et il en est de même du *film-de-Macbeth*.

Toutefois, si on branche les globes oculaires de Macbeth, qui sont activés, je le rappelle, par des fibres optiques capables d'être activées par la projection de rayon lumineux, sur le *film-de-Macbeth*, ils se mettront en mouvement deux fois avec les 13000 secondes d'écart. On peut faire plus. En filmant une version de Macbeth dans lesquelles les fibres optiques -correspondant aux neurones qui paralysent les sorties moteurs actionnant les cordes vocales- sont artificiellement détruites. Dans ce cas, si on branche le *film-de-Macbeth* aux cordes vocales de Macbeth, on entendra Hamlet parler dans son rêve lors du *film-de-Macbeth*. Ces témoignages ne sont sans doute pas convaincants, aussi passons à l'expérience suivante.

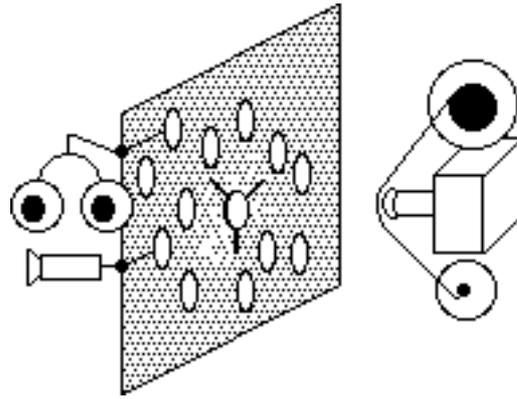
### 3°) deuxième expérience

Nous enlevons un module (un "&" par exemple) de Macbeth dans son état initial du début du rêve. Cette nouvelle version de Macbeth est appelée Macbeth<sup>-1</sup>. Macbeth<sup>-1</sup> n'est plus fonctionnellement équivalent à Macbeth. Qu'à cela ne tienne, nous lançons Macbeth<sup>-1</sup> et nous projetons le *film-de-Macbeth* obtenu tout à l'heure, en temps et espace (!) réel sur Macbeth<sup>-1</sup>.

Le schéma suivant schématise de façon un peu symbolique l'expérience :

---

<sup>6</sup> Dans la littérature la rage de dent fait souvent l'office présent du rêve. Elle aussi peut être jugée indépendante des entrées sensorielles (cf les douleurs aux membres fantômes), mais de façon moins explicite que le sommeil paradoxal. La critique de la théorie des rêves de Malcolm que je propose dans la section précédente justifie avec MDI le présent usage du rêve.

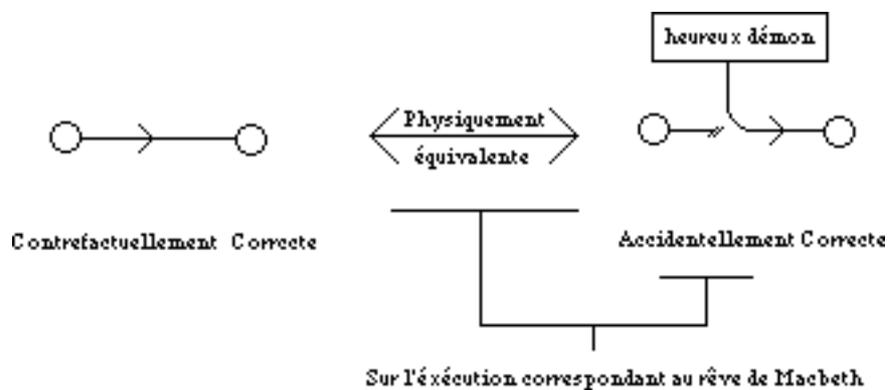


Supposons qu'en cours du procès deux rayons lumineux transportés par les fibres internes arrivent au "&" manquant. Celui-ci, étant absent, n'est pas à même de réaliser sa fonction, mais au même moment, parce que le film du graphe est projeté sur le graphe et qu'il y a correspondance lumineuse grâce au principe de détermination, le film enverra de façon appropriée le signal de sortie aux connecteurs du voisinage du "&" *défaillant*.

Aucun parmi ces connecteurs n'a jamais été supposé être à même de reconnaître l'**origine** de la source lumineuse. Le contraire signifierait que les modules sont capables de précognition, mais au niveau ou MEC-IND est adéquat les modules ne sont pas même supposés capables de cognition. En effet, comme dit Dennett 1978 une théorie de l'intelligence ne peut pas postuler de l'intelligence au départ dans ses éléments constitutifs, et c'est là une motivation philosophique pour MDI.

Le film joue donc le rôle, dans ce cas, de répertoire de démons-AC-*heureux-rayon* appropriés pour l'exécution particulière du rêve de Hamlet exécuté par Macbeth. Pour cette exécution particulière Macbeth<sup>-1</sup>, accompagné du film-de-Macbeth projeté sur lui-même, (dénommé Macbeth<sup>-1</sup>+*film-de-Macbeth*) est physiquement équivalent à l'activité de Macbeth.

Si on admet le principe de supervénience, nous devons admettre que le rêve de Hamlet supervène sur Macbeth<sup>-1</sup>+*film-de-Macbeth*.



Le même raisonnement vaut pour Macbeth dont on ôte d'abord deux modules, puis trois modules, etc. Si  $n$  = le nombre de modules de Macbeth, on a démontré que si le rêve supervène sur l'activité de Macbeth, il supervène sur  $\text{Macbeth}^{-n} + \text{film-de-Macbeth}$ . Mais  $\text{Macbeth}^{-n} + \text{film-de-Macbeth} = \text{film-de-Macbeth}$ . Donc, si le rêve de Hamlet supervène sur Macbeth, le rêve de Hamlet supervène sur le *film-de-Macbeth*.

Ainsi, si on admet la thèse de la supervénience physique, qui concerne des exécutions particulières de machines, on doit reconnaître que le mécanisme accidentel (avec heureux démons) véhicule les expériences subjectives correspondantes. En résumé :

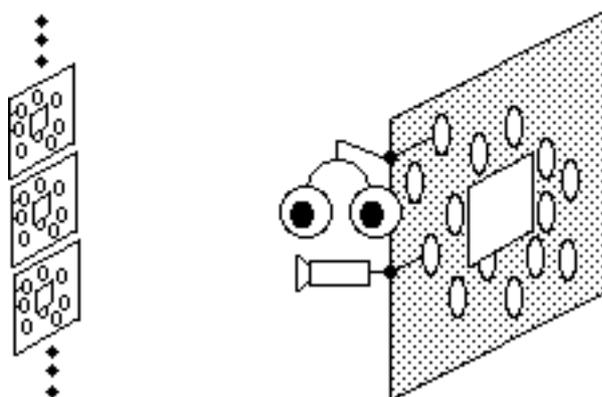
$$\text{SUP} + \text{MEC} \Rightarrow \text{MEC-AC-heureux-démon}$$

MEC est mis pour MEC-FORT ou MDI.

On pourrait s'arrêter ici puisqu'on aboutit à la proposition qu'aucun philosophe sensé ne peut admettre. Dans ce cas on déduit de suite qu'il faut abandonner l'hypothèse de supervénience physique ou l'hypothèse du mécanisme (en l'occurrence MEC-DIG-IND<sup>7</sup>). Pour ceux qui sont cependant prêt à admettre que le *film-de-Macbeth* véhicule l'expérience du rêve, essayons de poursuivre l'expérience afin de mettre en évidence une absurdité plus flagrante.

#### 4°) troisième expérience

On enlève une image d'un ou de plusieurs modules sur la pellicule même du film, et on enlève la partie correspondante de l'écran :



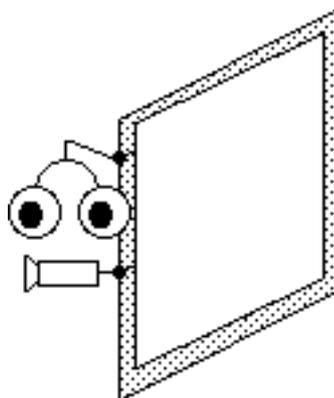
A gauche j'ai représenté une portion de la pellicule, à droite, se trouve l'écran du film de Macbeth, relié toujours aux globes oculaires et aux instruments vocaux (il ne s'agit **pas** du graphe booléen).

<sup>7</sup> Et même MEC-DIG-FORT, en vertu du principe de Webb ou de LWV local.

On projette ce qui est appelé naturellement le *film<sup>-1</sup>-de-Macbeth* sur l'écran dont il manque la partie correspondante. La conscience du rêve de Hamlet supervène-t-elle sur cette activité ?

Si l'on admet que le film-de-Macbeth véhicule l'expérience, alors il est difficile de justifier que l'expérience ne supervène pas au *film<sup>-1</sup>-de-Macbeth*. En effet le comportement au voisinage du trou n'est en rien modifié par la présence du trou qu'il soit dans la pellicule ou dans l'écran (ou les deux comme ici). Le trou, avec l'aide du projecteur, joue le rôle d'un heureux démon puisque ses "sorties" lumineuses sont identiques (accidentellement) aux sorties que l'on aurait vu sur le film sans trou, ou sur Macbeth lui-même. Les globes oculaires ne se mettent-ils pas encore en branle au moment opportun ?

En poursuivant un raisonnement toujours identique, il en est de même pour le *film<sup>-2</sup>-de-Macbeth* , ainsi que *film<sup>-3</sup>-de-Macbeth*, et même enfin, pour le *film<sup>-n</sup>-de-Macbeth* :



Comme le *film<sup>-n</sup>-de-Macbeth* est équivalent à l'absence d'activité (ou l'activité vide si on veut), admettre que le film-de-Macbeth véhicule (en temps *mesuré par Macbeth*) l'expérience, revient à admettre que l'activité vide véhicule (en temps mesuré par Macbeth) l'expérience. De plus ce raisonnement marche quel que soit le rêve. Donc tous les rêves supervènent (à tout moment, pour tous les temps mesurés par tout le monde) sur l'absence d'activité et çà, en temps-interne-de-Macbeth<sup>8</sup>, pour tous les Macbeth concevables.

### 5°) Conclusion

On a montré

SUP-PHYS + MEC => MEC-AC => absurdité

---

<sup>8</sup> Correspondant par exemple au temps interne défini par le S4Grz d'un Macbeth rêvant *lucidement*. Mais on peut prendre un temps interne à la Dewitt-Wheeler si on choisit un niveau de fonctionnalisme plus bas (pour une description concise de ce temps interne, voir Brout et Englert 1987).

si on tient cette absurdité pour une fausseté, on a montré :

$$\neg \text{MEC} \vee \neg \text{SUP-PHYS}$$

*Remarque et conjecture*

Ce raisonnement peut être formalisé dans une théorie auto-prouvablement  $\Sigma_1$ -complète. Une machine suffisamment introspective du type de celle décrite dans la deuxième partie (qui peut se *croire* universelle, et même le *savoir* localement) est capable d'aboutir "correctement" à cette conclusion. Pour nous, humain qui contemplons une telle machine, il va de soi (avec au moins MEC-DIG-FORT) qu'elle aurait "prouvé", à nos yeux :

$$\neg \text{SUP-PHYS}$$

puisque nous savons, dans ce cas-ci qu'il s'agit d'une machine. La machine elle-même pourrait déduire, *non correctement* :

$$\neg \text{MDI}$$

Bref, un nouveau parcours à travers la récursion et la logique modale nous amènerait à juger SUP-PHYS beaucoup moins plausible que MDI. Le paradoxe du graphe filmé peut subir le même sort que la réfutation de Lucas.

Avec S4Grz pour le connaissable et KD?/KD?\* pour la croyance, on devrait retrouver une application d'un principe de Lao Tseu-Watts-Valadier local<sup>9</sup>.

### 3.2.8 Comparaison avec Maudlin (1989)

Je présente à présent *Olympia*, la machine de Maudlin (1989).

Elle permet de donner une nouvelle démonstration de la proposition selon laquelle le mécanisme (a priori contrafactuel), en présence de l'hypothèse de supervénience physique, doit inclure l'accidentel, sous forme d'absence de pièce :

$$\text{MEC} + \text{SUP-PHYS} \Rightarrow \text{MEC-AC-absence-de pièce} \Rightarrow \text{absurdité}$$

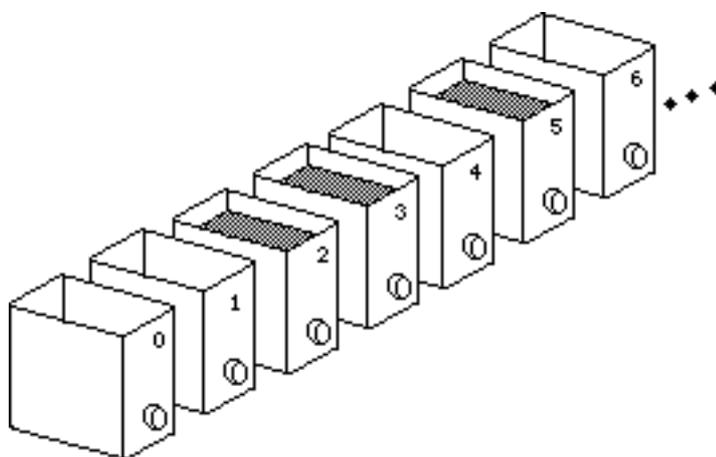
---

<sup>9</sup> La difficulté majeure d'une telle formalisation de *PGF* ou de *Maudlin* revient à trouver une interprétation arithmétique pertinente du (physiquement) *contrafactuel*. On verra en 3.3 comment cette difficulté est contournée en définissant le "physique" (le factuel) comme l'immédiatement inférable (ce qui n'est pas arithmétiquement trivial).

### 1°) Klara

Pour construire Olympia, il faut d'abord construire toute une collection de *Klara*. Klara est avant tout une machine universelle de Turing. Peu importe la nature de sa réalisation. On part de l'hypothèse MEC-FORT, ou MEC-IND. Maudlin suggère, pour son unité de calcul, un système mécanique comme une horloge (ou la machine de Babbage). Il impose cependant, pour accompagner la démonstration, une forme physique bien particulière au ruban de la machine. Le ruban de Klara est constitué d'une suite d'auges vides ou remplies d'eau.

Chacune de ces cuves est donc soit remplie (1 bit d'information) soit vide (0 bit d'information). La suite est potentiellement infinie, comme tous les rubans des machines universelles :



L'organe de lecture de Klara est un flotteur qu'elle transporte en parcourant le ruban, ou n'importe quel dispositif permettant de décider si une auge est remplie ou non. L'organe d'écriture est un robinet (ouvert) ou n'importe quel dispositif permettant de remplir une auge. Klara peut effacer le contenu d'un registre en vidant l'auge correspondant à ce registre.

Une particularité de l'unité de calcul de Klara est qu'il est possible de la bloquer dans un état, et de la débloquer ultérieurement de telle façon qu'elle continue normalement l'exécution qui avait été bloquée.

Klara, étant universelle, est à même, avec MEC-FORT (= MEC-FORT-CC) de véhiculer une expérience subjective. Pour fixer les idées, le programme considéré véhicule le rêve de Hamlet<sup>10</sup>, mais le raisonnement, ne nécessite pas MEC-IND. Klara peut émuler le programme correspondant au rêve de Hamlet en émulant le cerveau rêvant de Hamlet ou le graphe booléen décrit précédemment.

---

<sup>10</sup> Il est curieux que Maudlin donne de si beaux noms aux machines. Les noms choisis par Maudlin sont des allusions explicites au *marchand de sable* de Hoffmann. Pour le programme il se contente d'évoquer un impersonnel  $\pi$ , en outre le programme est censé véhiculer une expérience subjective indéfinie.

Notons que dans l'édition allemande, et dans la traduction française de M. F. Demet 1991, Olympia s'écrit avec un i : Olimpia et Klara avec un C : Clara. J'ai gardé l'orthographe de Maudlin.

Lorsque Klara procède le rêve de Hamlet depuis le stade initial jusqu'à la fin du rêve (ou un quelconque autre moment pourvu qu'il soit différent du stade initial), elle passe successivement par une suite d'états,

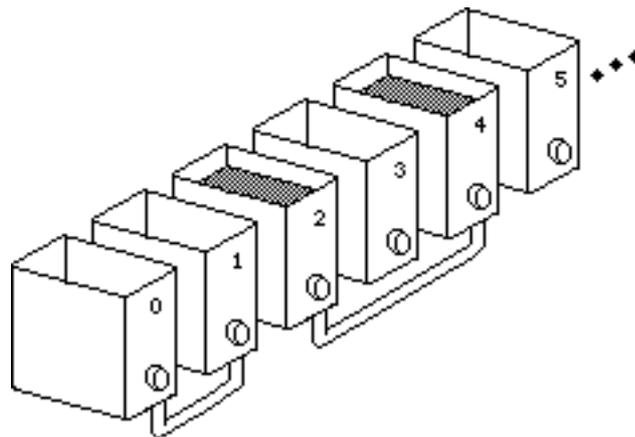
$$E_{(1)}, E_{(2)}, E_{(3)}, \dots, E_{(N)},$$

et son organe de lecture se déplace sur le ruban et visite ainsi successivement une suite d'auges,

$$A_{(1)}, A_{(2)}, \dots, A_{(N)}$$

correspondantes. Sa tête lit (avec le flotteur) le contenu du registre, le vide, le remplit ou le laisse en plan selon l'état  $E_{(i)}$ . Et l'état  $E_{(i)}$  détermine si en passant à l'état suivant,  $E_{(i+1)}$ , elle reste au-dessus de la même auge (auquel cas  $A_{(i)} = A_{(i+1)}$ ), elle se déplace sur l'auge de droite ou encore elle se déplace sur l'auge de gauche. Ces mouvements, actions et changements d'états sont déterminés par la table-programme de Klara qui implémente ainsi une machine universelle de Turing.

Pour construire Olympia à partir de Klara, on aimerait encore pouvoir allouer un registre à plusieurs endroits (à plusieurs adresses du ruban). On exige que le contenu d'un registre (d'une cuve) du ruban de Klara puisse être alloué sur plusieurs adresses. Vider ou remplir un tel registre aura pour effet de vider ou remplir les cuves aux adresses allouées. A cet effet des tubes de communication peuvent être établis entre des registres contigus sur base du principe des vases communicants :



#### *Dernière modification de Klara*

Sans rien changer à l'activité potentielle et universelle de Klara, on peut réorganiser son ruban de telle façon que chacune des auges est visitée une fois et dans l'ordre du ruban :

$A(i) = \text{l'auge } i$

Elle visite l'auge 0, puis l'auge 1, etc. Il suffit pour ça de coder une transformation des adresses du ruban dans la table-programme de Klara. Dans une situation où elle doit retourner en arrière selon la table-programme de Klara avant la réorganisation du ruban, on bialoue le registre suivant avec le précédant, si elle doit rester sur place on bialoue le registre courant avec le suivant. Parce que je ne permets que des déplacements d'une unité chez Klara, les bi-allocations du type aperçu dans la figure précédente suffiront à la tâche<sup>11</sup>.

A présent, avec cette transformation, lorsque Klara émule le programme correspondant au rêve de Hamlet (et donc véhicule la conscience et, par exemple, la lucidité du rêve, avec SUP-PHYS), elle déplace sa tête sur son ruban séquentiellement de l'auge 1 à une auge  $n$  déterminée ( $n = N$ ), en remplissant, en vidant ou ne modifiant pas, chacune de ces auges.

### 2°) Proto-Olympia

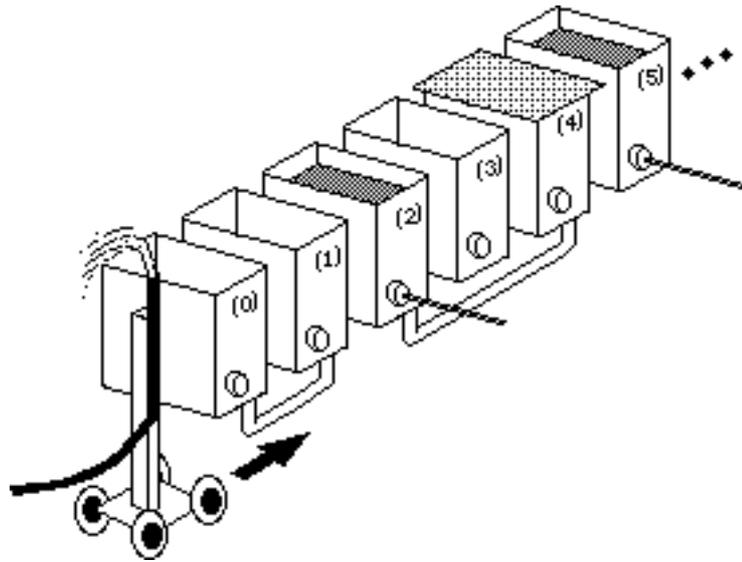
Si Klara joue le rôle de Macbeth, proto-Olympia joue le rôle du *film-de-Macbeth* restreint aux changements du ruban. Proto-Olympia possède le même ruban que Klara lorsque celle-ci véhicule le rêve de Hamlet, et le modifie de la même façon que Klara en passant elle-même par une suite de  $N$  états distincts.

Proto-Olympia commet cependant ces modifications de façons paresseuses, et minimise le travail au maximum. De plus ces modifications sont accidentelles, pour ne pas dire *démoniaque*.

Proto-Olympia est constitué d'un dispositif sur roues se déplaçant le long d'un ruban identique à celui de Klara émulant le rêve. Elle remplit, vide ou ne modifie pas les auges de la même façon que Klara par l'intermédiaire d'artifices disposé à même le ruban :

---

<sup>11</sup> La tête de la machine originale de Maudlin peut faire des sauts arbitrairement grands sur son ruban, ce qui nécessite une tuyauterie plus sophistiquée. Ces détails ne sont pas pertinents puisqu'il existe des machines de Turing universelles de chaque sorte. On aurait aussi pu prendre une machine universelle de Turing qui n'efface jamais le ruban (pour l'existence d'une telle machine, voir Wang 1957, Minski 1967).



Dans la situation décrite sur la figure, proto-Olympia va remplir le premier registre A(0), donc le second A(1), vider A(2) en frappant la baguette reliée au bouchon, donc vider A(4) par bi-allocation, remplir A(3), ne plus remplir A(4), vider A(5) etc.

Le rêve de Hamlet supervène-t-il sur l'activité de proto-Olympia ? La question semble franchement absurde à ce stade. Proto-Olympia n'a pratiquement aucune activité computationnelle, et néglige toute l'activité interne de Klara. De plus, si on change le contenu du ruban, le comportement de proto-Olympia ne reflétera plus celui de Klara. La manipulation du ruban par proto-Olympia était correcte pour des raisons purement accidentelles.

En bref, proto-Olympia est équivalente à Klara pour l'exécution particulière correspondant au rêve de Hamlet et cela n'est vrai, en plus, que si l'on se restreint à la manipulation du ruban.

La seule raison que l'on pourrait invoquer à présent pour prétendre que proto-Olympia véhicule le rêve de Hamlet reposerait sur le paradoxe du graphe filmé où l'on aurait admis la conclusion selon laquelle le vide véhicule toutes les expériences (et pourquoi dès lors proto-Olympia ne véhiculerait-elle pas le rêve ?), mais en ce qui concerne le raisonnement de Maudlin, ce serait mettre la charrue avant les boeufs.

#### 4°) Olympia

Le passage de Klara à proto-Olympia est le passage d'une machine contrefactuellement correcte à une machine paresseuse (peu d'activité physique) accidentellement correcte (relativement à la gestion du ruban).

Le passage de proto-Olympia à Olympia est le passage d'une machine paresseuse accidentellement correcte à une machine contrefactuellement correcte, mais tout autant paresseuse.

Olympia continuera à manipuler le ruban de façon paresseuse et accidentelle mais seulement *pour l'exécution particulière* du rêve. Si le ruban est soudainement altéré, Olympia va se mettre dans une grande activité en tout point comparable à celle de Klara, de Macbeth ou du cerveau de Hamlet.

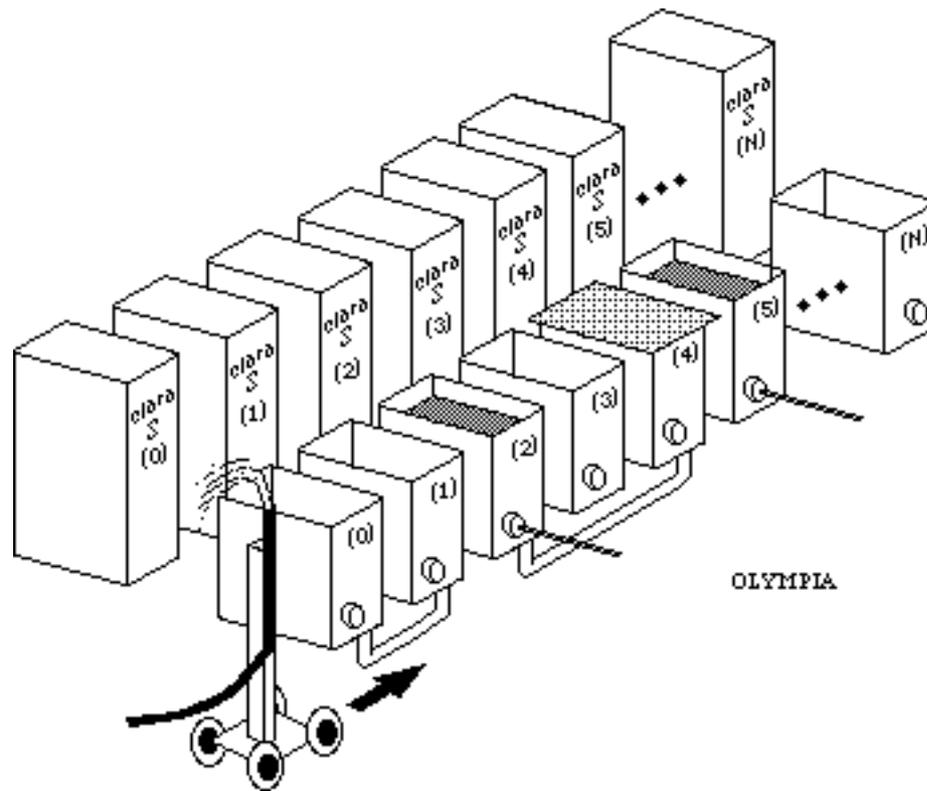
Pour ajouter la possibilité d'un comportement contrefactuellement correct, Maudlin ajoute à proto-Olympia une description de tous les états possibles qu'elles aurait pu rencontrer si le ruban avait été différent<sup>12</sup>, c'est-à-dire, ces états qui auraient été atteints par le programme du rêve de Hamlet, l'eussions-nous perturbé.

L'idée de Maudlin est d'attacher à chaque registre du ruban de proto-Olympia une version *bloquée* de Klara (rappelons-nous que le blocage est une caractéristique de Klara). De plus chaque Klara est bloquée dans l'état E(i) correspondant au registre A(i). Je rappelle que les états E(i) correspondent à l'exécution normale du rêve.

Il ne reste plus alors qu'à placer dans les registres A(1) ... A(N) du ruban d'Olympia un dispositif (avec loquet et flotteur par exemple) tel que si Olympia rencontre un registre A(k) au contenu différent du A(k) de l'exécution du rêve par Klara, la partie proto-Olympia d'Olympia débloque la Klara afférente à ce registre. Comme celle-ci est dans l'état E(k) elle va continuer l'exécution correctement corrigée du rêve. Olympia est ainsi contrefactuellement correcte :

---

<sup>12</sup> Un ensemble d'états accessibles par une machine digitale peut être représenté de façon finie au moyen (du code) d'une machine universelle finie, comme Klara (avec la thèse de Church).



Le dispositif débloquant Klara (non représenté sur le dessin) est actif seulement dans la situation contrefactuelle. Autrement dit, si le ruban correspond au ruban de Klara émulant le programme du rêve, Olympia "émule le programme du rêve" sans jamais activer aucune des Klara. Il est, avec MEC, correct de dire qu'Olympia véhicule le rêve puisqu'Olympia est contrefactuellement correcte. D'autre part vu l'inactivité physique des Klaras, l'activité d'Olympia est physiquement équivalente à celle de proto-Olympia sur l'exécution particulière du rêve de Hamlet.

Avec l'hypothèse de supervénience physique SUP-PHYS, nous devons conclure, que proto-Olympia, dont l'activité est strictement identique à celle d'Olympia avec la suppression des pièces inactives (MEC-AC-*absence*), véhicule l'expérience véhiculée par Olympia.

Maudlin s'arrête ici considérant que l'absurdité est atteinte. Il n'essaye pas de déshabiller davantage proto-Olympia pour arriver au vide, mais cela n'est plus très difficile (voir plus loin).

### 5°) Conclusion

Parce que Maudlin tient au principe de supervénience physique et à l'hypothèse du matérialisme, il suggère l'abandon du mécanisme (Maudlin 1989). La preuve de Maudlin admet les mêmes conséquences que le paradoxe du graphe filmé. Maudlin et PGF montrent que MEC + SUP-PHYS conduit à une absurdité. Parce que nous avons adopté MEC depuis le départ, il est naturel d'abandonner plutôt l'hypothèse de supervénience physique. Comme la motivation pour celle-ci résidait dans la tentative de concilier le

matérialisme avec le mécanisme, c'est le matérialisme naïf<sup>13</sup> qu'il conviendra de rejeter.

*Remarque*

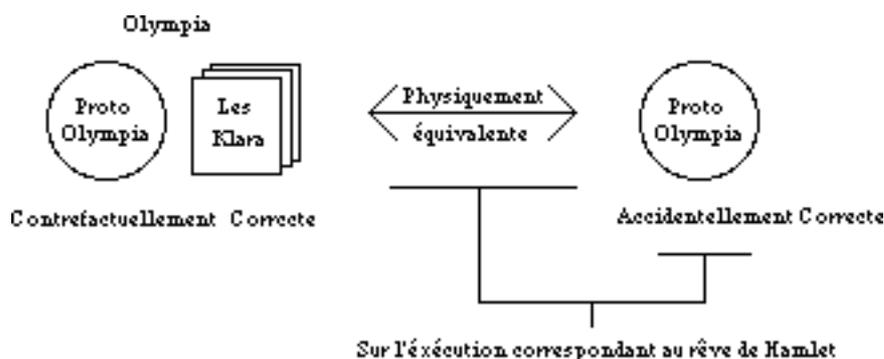
Il ne faut pas se laisser distraire par le caractère irréaliste, et sans doute même un peu *baroque*, des machines considérées dans cette section. Elles nous ont permis de montrer que le mécanisme fort (et donc en particulier MDI) est incompatible avec la thèse de la supervénience physique. Elle nous oblige à admettre que, ce que Malcolm affirme au sujet des rêves (leur non supervénience physique) s'applique, non pas *possiblement* comme le suggère déjà les paradoxes de la duplication, mais *nécessairement*, à la réalité ou aux états de veille<sup>14</sup>.

### 3.2.9 Macbeth interprète Olympia

Ce point est tellement contre-intuitif qu'il n'est pas trop luxueux de s'apesantir un peu plus. Je propose de retraduire Maudlin en terme de graphe filmé.

L'originalité de Maudlin par rapport au graphe filmé est située dans le fait qu'il construit en deux étapes une machine contrefactuellement correcte physiquement équivalente, pour une exécution, à une machine qui n'a pratiquement aucune activité physique. Les deux étapes sont constituées du passage de Klara à proto-Olympia, et du passage de proto-Olympia à Olympia.

Les Klara jouent le rôle de démon-absence-de pièces pour proto-Olympia :



Avec PGF, les deux étapes sont mélangées dans le premier déshabillage de Macbeth et le principe de supervénience physique doit être appelé à chaque étape du déshabillage, si bien que l'on passe de façon presque

<sup>13</sup>J'oppose le matérialisme naïf à d'éventuelles interprétations non matérialistes de la physique (quantique). Le mécanisme n'exclut pas une forme de physicalisme. En 3.3, je montre qu'au contraire un minimum de physicalisme doit pouvoir être extrait, par MDI, de l'arithmétique du dehors.

<sup>14</sup>Je n'ai toujours pas introduit d'ontologie, état et monde sont toujours interchangeables.

continue de la machine contrefactuellement correcte (Macbeth) à la "machine" accidentellement correcte (film-de-Macbeth). On pourrait dès lors objecter que la conscience se dilue progressivement lors du déshabillage<sup>15</sup>.

L'avantage de l'approche de Maudlin est de transformer la suite progressive des substitutions en l'enlèvement, d'un coup, d'une partie physique clairement inactive (dans une exécution particulière) d'une machine contrefactuellement correcte (Olympia) donnant directement une machine purement accidentellement correcte (proto-Olympia).

Je propose donc d'explicitier cette caractéristique de la preuve de Maudlin en utilisant des modifications de Macbeth. Je réexaminerai de même le second déshabillage.

Macbeth est un graphe booléen planaire. Il est capable (avec MEC-FORT) de véhiculer le rêve de Hamlet. De plus cette émulation est, comme celle de Klara, contrefactuellement correcte, grâce à quoi il est possible de réveiller Hamlet. Je suppose qu'on a ajouté les entrées sensorielles de l'ouïe au graphe, si bien qu'il est possible, en criant par exemple, de perturber le rêve ou de réveiller Hamlet.

Macbeth est donc bien placé pour jouer le rôle de Klara.

Le film-de-Macbeth est purement accidentellement correcte. Impossible de réveiller Hamlet. Il correspond naturellement à proto-Olympia qui est incapable de tenir compte d'un changement contrefactuel du ruban.

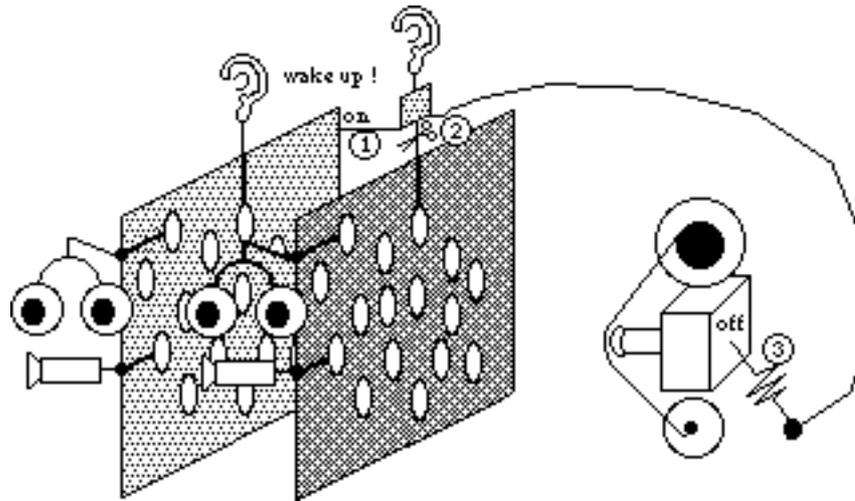
Olympia est née d'un mariage sophistiqué entre proto-Olympia et Klara. proto-Olympia gère l'exécution accidentelle du rêve, Klara, multipliée, gère inactivement, le contrefactuel.

Il suffira donc de marier Macbeth et le film-de-Macbeth de façon analogue. Je rappelle que dans les figures,  représente Macbeth, et que  représente le film-de Macbeth.

Voici la nouvelle *Olympia* :

---

<sup>15</sup> Comme dans le paradoxe du petit tas de sable qui se transforme "continuellement" en grand tas de sable, par addition de grain de sable (sorite).



Le film du graphe, la nouvelle proto-Olympia, est placée entre Macbeth dans un état inactif.

Il a fallu lui adjoindre un appendice booléen relié à un dispositif sensoriel sonore, ainsi qu'à une paire de ciseaux.

On peut exiger de ce dispositif lui-même qu'il ne se mette en branle qu'à condition qu'un son d'une assez haute intensité soit émis dans sa proximité.

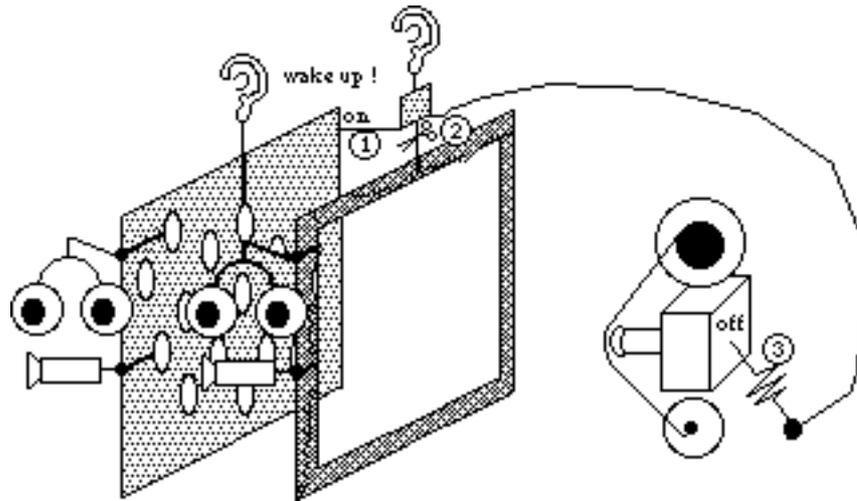
Dans ce cas-là l'appendice booléen s'active. Il gère le signal sonore, envoie la réponse vers le film-de-macbeth, mais surtout, il active Macbeth, il coupe avec les ciseaux le film et laisse actif un cours instant le projecteur afin de transférer l'état instantané du film-de-Macbeth sur Macbeth, l'initialisant ainsi dans l'état (correct) du rêve au moment où le signal sonore a été émis. Comme proto-Olympia qui s'en remet à Klara (*aux Klara*) face à une perturbation, le dispositif présenté permet au film-de Macbeth d'activer Macbeth lui-même dans une situation similaire.

Comme Olympia, la nouvelle Olympia est contrefactuellement correcte.

Pour l'exécution du rêve de Hamlet la nouvelle Olympia est physiquement équivalente au film de Macbeth (au dispositif d'alerte de l'appendice Booléen près). Le principe de supervénience physique nous permet alors de retirer les parties contrefactuelles inactives, et nous force à conclure que le film-de Macbeth véhicule le rêve. Absurde. Donc le mécanisme ou la supervénience sont faux.

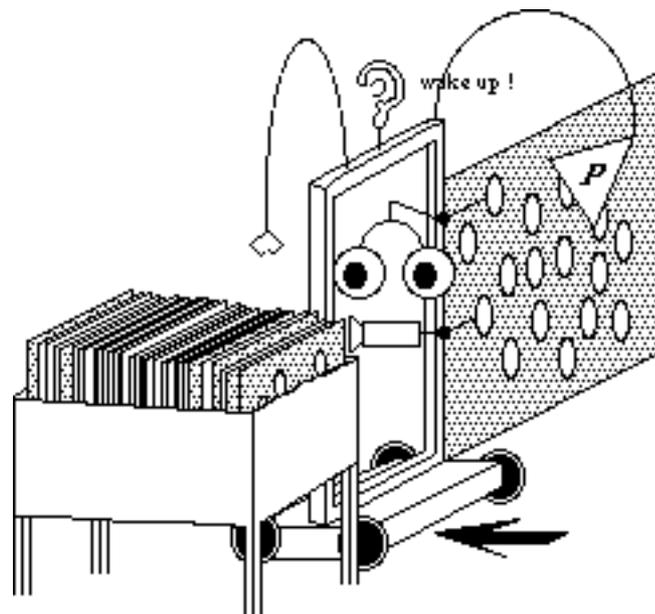
Absurde ? La nouvelle Olympia ne donne-t-elle pas un argument pour admettre que le premier déshabillage du graphe filmé ne supprime pas l'expérience de la conscience tout en interdisant le second déshabillage ?

L'argument suivant *semble* en effet distinguer le film-de-Macbeth du film vide. Remplaçons le film-de-Macbeth par le film vide dans la nouvelle Olympia :



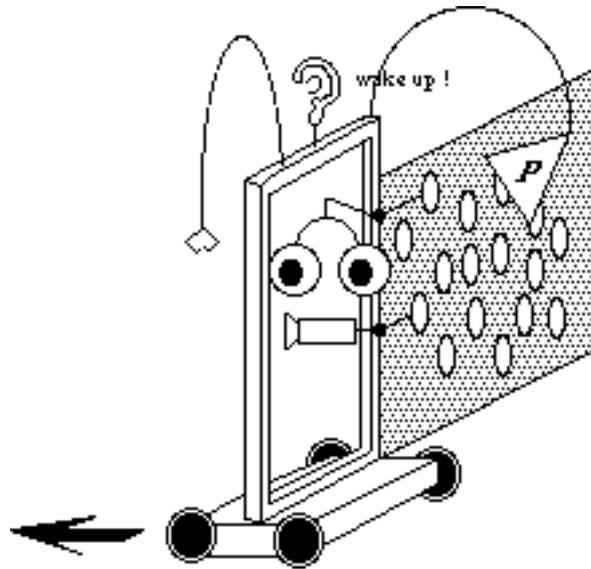
Rappelons nous que le "trou" est présent dans la pellicule également. Dans ce cas il n'y a aucune chance que Macbeth récupère un quelconque état instantané du procès du rêve.

Cet argument n'est pas pertinent. La nouvelle Olympia n'a pas été conçue pour marier le *film vide* et Macbeth. En particulier le dispositif suivant, 3ème Olympia, pallie à cet inconvénient :

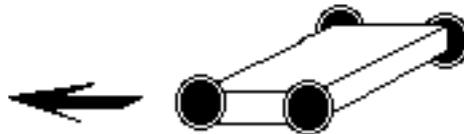


Il s'agit de Macbeth éteint, se déplaçant le long d'une boîte comprenant sous forme de diapositives les états successifs du film-de-Macbeth. Le dispositif est muni d'un système d'alerte sonore. Si on crie "wake up", il sélectionne au moyen de la grue la diapositive en face de lui et au moyen du projecteur p, initialise Macbeth en l'activant simultanément. Cette nouvelle Olympia (transportant Klara *silencieuse* sur son dos) est contrefactuellement correcte.

Le "proto-olympia" correspondant est obtenu en retirant ce qui est physiquement inactif. On retire dès lors la boîte à diapositive :



et puisque Klara-Macbeth est éteinte, le dispositif est encore physiquement équivalent<sup>16</sup> à :



Une planche à roulette véhicule-t-elle le rêve de Hamlet ?

L'activité de la planche à roulette a en commun avec l'activité vide de ne pas avoir de relations privilégiées avec Hamlet.

Si la machine Macbeth ou la machine Klara véhicule le rêve de Hamlet (en temps mesuré par Macbeth), **alors** le rêve supervène à l'activité de chacune des Olympia. Si la conscience supervène sur l'activité **physique** du dispositif computationnel, alors la conscience supervène sur l'activité des proto-Olympia incluant le vide et la planche à roulette. Absurde.

### 3.2.10 Conclusion

Si on veut conserver l'hypothèse mécaniste digitale et indexicale, MEC-DIG-IND, on doit abandonner le principe de la supervénience de la conscience sur l'activité **physique** du dispositif computationnel.

Si, conservant l'hypothèse du mécanisme, on admet que la conscience supervène encore sur quelque chose attaché au dispositif computationnel,

---

<sup>16</sup> en ce qui concerne l'activité computationnelle. Si on abandonne le mécanisme, on peut réintroduire une supervénience physique qui tienne compte de la présence de telle ou telle partie "matérielle" arbitraire.

elle supervène alors, en absence de SUP-PHYS, sur la structure des états computationnels<sup>17</sup>.

Si un onirophysiologue, ou un psychanalyste, désire étudier le rêve de Hamlet, clairement il a plus de chance d'apprendre quelque chose au moyen du film-de-Macbeth, ou de la boîte de diapositives que de la planche à roulette ou du vide. S'il dispose en plus de Macbeth, il pourra non seulement étudier le rêve, mais écouter les commentaires de Hamlet lui-même. Ceci est déjà une conséquence de MEC-BEH.

Chaque diapositive, comme chaque morceau suffisamment grand de la pellicule du film de l'ordinateur est équivalent, relativement à Macbeth, au cristal de la duplication postposée. L'apparence concrète et singulière du dispositif computationnel est une construction indexicale permettant à la conscience de se manifester relativement à un environnement localement universel.

D'ou vient cette apparence concrète ? A la fin de son article, Maudlin reconnaît une analogie entre Olympia et le cinéma :

*Olympia creates the illusion of being Klara in action in much the same way that a movie creates the illusion of motion on the screen. Maudlin (1989)*

Il conclut

*The problem with computationalism is that it does not contain the conceptual resources to distinguish the flickering illusion from reality. (idem)*

Mais ici encore la question est de savoir d'où une telle "flickering" illusion peut-elle provenir. Dans l'expérience par la pensée du graphe filmé, on a supposé implicitement qu'Hamlet a du commettre le rêve "réel" au moins une fois, et vivre les expériences antérieures nécessaires à la construction nocturne du rêve.

Un tel état instantané peut avoir un coût computationnel élevé, il peut s'agir d'un état relativement rare. Comme l'oeuvre de Proust "unique" et duplicable, l'état du cerveau de Hamlet (ou de son corps, ou de son univers, etc.) pourrait être computationnellement et intrinsèquement rare. La thèse de Church peut rendre de telles notions machine-ou-formalisme-indépendantes.

Le film du graphe joue le rôle d'une suite de cristal ayant capturer une suite d'états instantanés. MEC, comme on l'a vu dans la deuxième partie, impose des contraintes non triviales sur les suites de tels états.

MEC-FORT impose que la possible *manifestation*, relative à un environnement, d'une conscience, nécessite un code et une machine

---

<sup>17</sup> Maudlin, gardant SUP-PHYS, abandonne MEC-FORT (et donc MDI). Comme il accepte l'idée que le cerveau a, au moins, un rôle computationnel, la conscience doit supervenir sur *plus* que la physicalité et le computationnel. Du point de vue mécaniste cela complique les choses puisqu'il faudra à présent expliquer les relations entre le computationnel, le physique, et ce plus. On retrouve les difficultés de la position dualiste (voir section suivante). Avec MEC la conscience supervène sur *moins* que cette *activité physique*.

universelle. Mais si l'on identifie l'état de conscience (indexical) à l'état d'une machine *physique* parcourant une exécution particulière, ce qui est nécessaire pour la thèse de l'identité ou pour le principe de supervénience physique, on pourra toujours reconstruire une machine du style Olympia et donc un paradoxe du genre du graphe filmé.

La conscience *vécue* devrait plutôt supervenir sur *l'existence* d'une suite de nombres correspondant aux états instantanés d'une machine autoréférentiellement et inférentiellement correcte relativement à un environnement universel local (et ses variétés contrefactuelles).

Une telle suite est nécessairement bifurquante, et il va falloir tenir compte de *toutes* les extensions consistantes *immédiatement* accessibles<sup>18</sup>.

Abandonner SUP-PHYS, nécessite, pour justifier le rôle relatif (dans ses manifestations) de la conscience, de rendre compte de l'apparence intersubjective de la matière, par exemple sous la forme de cohérence entre des extensions consistantes.

C'est de cette façon que le mécanisme impose de construire une phénoménologie de la matière à partir de considérations computationnelles ou immatérielles plutôt que l'inverse.

Je tente de préciser ce point dans la dernière section. Elle nécessite de rendre les considérations philosophiques plus explicites, notamment au sujet de la nature de l'engagement ontologique minimal nécessaire pour poser le problème.

Ce que Maudlin néglige, c'est que la thèse de Church, et a fortiori l'utilisation d'éventuels stratagèmes intensionnels, fait de la collection des états computationnels possibles une structure non triviale possédant d'éventuels champs d'états rares et corrélés. Pas seulement ça : les discours des machines elles-mêmes, dans les voisinages de l'infini, sur ces états immédiatement accessibles, sont aussi non triviaux.

Maudlin néglige la possibilité de SUP-COMP.

PGF et cette conception purement computationnelle de l'esprit soulèvent a priori un nouveau paradoxe dû au "trop-plein" a priori de ces états possibles : le paradoxe du dovetelleur universel (PDU).

PDU permettra de se faire une idée de la non-trivialité de la structure des états directement accessibles et des discours limites des machines plongées dans cette structure au sujet de cette structure.

Il s'agira d'un *embryon de géométrie des rêves*, mais aussi d'un *décor* arithmétique pour une reformulation mécaniste du problème du corps et de l'esprit.

---

<sup>18</sup> Il serait plus approprié de parler d'un *nuage* de nombres, puisque cette suite peut être bifurquante (et même fusionnante si on pense aux états multiples d'Everett). revoir 1.3, 2.3 et voir 3.3.

## Bibliographie locale

**BROUT R. & ENGLERT F., 1987**, *Cosmologie Quantique, Aux confins de l'univers*, Fayard/Fondation Diderot, pp. 269-289.

**DEMET M.-F., 1991**, Traduction de Don Juan, et Der Sandmann, de E. T. A. **HOFFMANN**, Le Livre de Poche 8738, Librairie Générale Française, Paris.

**DENNETT D. C., 1978**, *Brainstorms*, Harvester Press, Hassocks, Sussex, 1979.

**DEWDNEY A. K., 1990**, *The Magic Machine. A Handbook of Computer Sorcery*, W. H. Freeman and Company, New-York.

**DENNETT D.C. and HOFSTADTER D., 1981**, (composed and arranged by), *Mind's I*, Basic Books, Inc., Publishers, New-York. (Paru en français sous le titre *vue de l'esprit*, interEditions, Paris, 1987).

**JOHNSON-LAIRD P., 1987**, *How Could Consciousness Arise from the Computations of the Brain ?*, in Mindwaves, C. Blakemore and S. Greenfield, Basil Blackwell, Oxford and New York, pp. 247-257.

**KRIPKE S., 1972**, *Naming and Necessity*, in Semantics of Natural Language, D. Davidson et G. Harman (eds), Reidel, Dordrecht, aussi Harvard University Press 1980, Trad. Franç. de Pierre Jacob et François Recanati : La logique des noms propres, Les Éditions de Minuit, 1982, Paris.

**LYCAN W. G., 1987**, *Consciousness*, The MIT Press, Cambridge.

**MACKAY D. M., 1960**, *On the Logical Indeterminacy of a Free Choice*, *Mind* 69, n° 273, pp. 31-40.

**MARCHAL B., 1988**, *Informatique théorique et philosophie de l'esprit*. Actes du 3ème colloque international de l'ARC, Toulouse .

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

**MINSKI M., 1967**, *Computation: Finite and Infinite Machines*, Prentice Hall.

**PUTNAM H., 1960**, *Minds and Machines*, *Dimensions of Mind : A Symposium*, Sidney Hook (Ed.), New-York University Press, New-York. Repris dans Anderson A. R. (Ed.), 1964.

**SEARLE J.R., 1980**, *Minds, Brains and Programs* dans *The Behavioral and Brain Sciences*, vol 3, Cambridge University Press. (repris dans DENNETT D.C. and HOFSTADTER D. 1981).

**SEARLE J.R., 1984**, *Minds, Brains and Sciences*. British Broadcasting Corporation. Trad. Franç. chez Hermann, 1985.

**WANG H., 1957**, *A variant to Turing's theory of computing machines*, *J of the ACM*, 4, 1.