

Jean LASSEGUE\*

## **La méthode expérimentale, la modélisation informatique et l'intelligence artificielle**

La notion d'expérience est riche d'une longue histoire qui tourne autour de trois notions-clés : le raisonnement hypothético-déductif, la mathématisation et l'instrumentation. L'intelligence artificielle, dont on a défini au préalable les deux grandes tendances ("science des conditions générales et effectives de l'intelligence" et "science des artefacts interactifs"), hérite de ce contexte mais en innovant grâce à la modélisation informatique : celle-ci instaure un nouveau rapport entre l'objectif et l'empirique. En réinterprétant la notion d'expérience, l'intelligence artificielle a mis au jour une nouvelle signification de ce terme, l'expérience de pensée, que deux notions nouvelles rendent possible : celle de jeu d'écriture et celle d'inconscient procédural.

*Mots-clés* : raisonnement hypothético-déductif, mathématisation, instrumentation, modélisation informatique, expérience de pensée, jeu d'écriture, inconscient procédural.

**Experimental method, computing modelisation and artificial intelligence.** The notion of experience has a long history which is focused on three key-notions: deductive reasoning, mathematisation and instrumentation. Artificial intelligence, defined either as the "science of the general and effective conditions of intelligence" or as the "science of interactive artefacts", inherits this history but goes a step further thanks to computing modelisation : from this point of view, a new relation between objectivity and empiricity is made possible. In reinterpreting the notion of experience, AI gives a new meaning to the idea of thought experiment. This idea is made possible thanks to two notions: that of writing procedure and that of procedural unconscious.

---

\* CNRS URA 1234. Université de Caen / E.N.S. Paris, 1 rue Maurice Arnoux, F-92120 Montrouge; lassegue@canoe.ens.fr.

Ce texte a été rédigé à partir d'une intervention à la journée de la MRSH de Caen, "Expérimentation en intelligence artificielle", le 9 juin 1995.

*Key Words* : deductive reasoning, mathematisation, instrumentation, computing modelisation, thought experiment, writing procedure, procedural unconscious

## 1. INTRODUCTION

Mon but est de montrer en quel sens l'intelligence artificielle a transformé notre interprétation de la notion d'expérience. Pour ce faire, je vais essayer de replacer la notion d'expérience en intelligence artificielle dans un contexte historique et épistémologique plus large. Historique : je vais faire intervenir des données provenant de l'histoire des sciences pour décrire grossièrement la notion d'expérience. Épistémologique : je vais essayer de mettre au jour un certain nombre de traits, que j'espère caractéristiques, touchant la question du statut de l'expérience quand on la met en rapport avec l'intelligence artificielle.

### 1.1. Le sens donné à l'expression controversée d'intelligence artificielle

Mais avant, il me faut préciser rapidement comment j'entends l'expression d'"intelligence artificielle".

Il est sans doute difficile de caractériser simplement une discipline comme l'intelligence artificielle parce que son appellation même apparaît au moins autant comme un slogan publicitaire que comme une prise de position théorique. Concernant cette difficulté, on invoque généralement l'histoire encore trop brève de la discipline, qui ne permettrait pas d'en décrire la nature. Mais cela me semble être une mauvaise raison : personne n'aurait par exemple du mal à définir une science comme l'informatique théorique dont l'histoire n'est pourtant guère plus longue. C'est donc plutôt que le champ d'investigation de l'intelligence artificielle n'est pas aussi stabilisé que ce que l'on pourrait croire et que le débat sur la nature, le but et les réalisations de l'intelligence artificielle ne produit pas encore d'unanimité. Ainsi contraint par l'état instable de la discipline, je me reposerai ici sur le travail d'interprétation réalisé par Y.-M. Visetti qui, dans plusieurs articles<sup>1</sup>, a effectué un repérage en synchronie et en diachronie des grandes tendances du domaine, des aspirations de ceux qui y travaillent ainsi que des liens que l'intelligence artificielle peut entretenir avec les autres champs du savoir, que ce soient les sciences "dures" comme les mathématiques et la physique ou les sciences "humaines" comme la linguistique ou la psychologie.

En amont de l'intelligence artificielle toutes tendances confondues, une science tutélaire : l'informatique théorique. Constituée en champ autonome à l'articulation des mathématiques discrètes et de la théorie logique de la calculabilité, l'informatique théorique a sa vie propre, encore trop négligée par les épistémologues et les philosophes des

---

<sup>1</sup> En particulier, Y.-M. Visetti (1989), (1991), (1992).

sciences<sup>2</sup>. C'est sans doute à cause de ce rôle tutélaire que l'on a pris — un temps — l'habitude de diviser l'intelligence artificielle en deux grandes tendances — elles-mêmes héritées de l'histoire de l'informatique théorique — selon que le calcul qui y est opéré s'effectue selon la norme classique de la théorie de la calculabilité (la machine de Turing) ou selon celle des réseaux de neurones. En fait, du point de vue de l'intelligence artificielle, cette différence est souvent secondaire parce que ce n'est pas au niveau de l'effectivité du calcul qu'elle se place<sup>3</sup>.

L'intelligence artificielle proprement dite se décline de deux façons, selon qu'on l'interprète comme "science des conditions générales et effectives de l'intelligence" ou comme "science des artefacts interactifs" et notamment des artefacts sémiotiques. Cette différence reflète aussi un déroulement historique puisque le second point de vue paraît être né des déconvenues liées au premier.

Dans le cas où l'on s'accorde à lui reconnaître le statut d'une "science des conditions générales et effectives de l'intelligence", — au même titre que la psychologie et sans chercher à la remplacer — l'intelligence artificielle est l'étude, par des moyens informatiques, du comportement intelligent sous ses formes naturelles ou artificielles<sup>4</sup>.

---

<sup>2</sup> Il s'agit pourtant d'un bel exemple de naissance d'une science. A noter cependant la thèse de philosophie de P. Wagner qui porte précisément sur la question du rapport entre l'informatique théorique et l'intelligence artificielle. Cf. P. Wagner (1994).

<sup>3</sup> Comme le dossier d'*Intellectica* consacré aux Fonctionnalismes dirigé par E. Pacherie l'a clairement établi, les réseaux de neurones formels restent en fait, et dans la majeure partie de leurs emplois, dans l'orbite du fonctionnalisme. Par ailleurs, si l'on prend par exemple l'article de Turing de 1950, fondateur pour l'intelligence artificielle classique, on voit que dans le jeu de l'imitation, c'est l'interaction avec la machine qui fait le fond du problème de l'intelligence et de sa simulation et non pas du tout de savoir quel type de machine — et partant quel type de calcul — intervient dans le jeu. Cf. A. M. Turing, (1950) : 433-460.

<sup>4</sup> C'est en fait la "vieille" définition de Von Neumann, proposée avant que l'expression "d'intelligence artificielle" ne devienne populaire et qui reste proche de la cybernétique de Wiener. Comme le rapporte Burks dans l'introduction au livre de J. Von Neumann, (1966) : 18 : «Comme il [Von Neumann] était conscient des importantes similarités existant entre les ordinateurs et les organismes naturels, ainsi que des avantages heuristiques à comparer des systèmes aussi différents mais apparentés, il rechercha une théorie qui couvrirait leur champ à tous les deux. Il appela la théorie qu'il proposait "théorie des automates". Cette théorie des automates devait être un corps de concepts et de principes cohérent touchant la structure et l'organisation des systèmes naturels et artificiels, le rôle du langage et de l'information dans ces systèmes ainsi que leur programmation et leur contrôle.»

Cette interprétation fait de l'action et de la pensée les deux faces d'une même médaille computationnaliste et mentaliste et l'intelligence artificielle apparaît dès lors comme la colonne vertébrale des sciences cognitives : elle leur apporte le secours de ses méthodes et les conditions de possibilité de ses modélisations<sup>5</sup>. Globalement fonctionnaliste<sup>6</sup>, l'intelligence artificielle aurait pour but ultime la constitution d'une psychologie scientifique de l'intentionnalité, qu'elle soit naturelle ou artificielle, fondée sur le concept de traitement de l'information : "l'esprit" interprété comme intériorité calculante serait ainsi au cœur des préoccupations de l'intelligence artificielle.

Dans le cas où l'on s'accorde au contraire à lui reconnaître le statut de "science des artefacts interactifs", représentatifs ou sémiotiques, l'intelligence artificielle ne viserait pas à fonder une psychologie scientifique de l'intentionnalité et n'aurait plus dès lors le rôle "directeur" qu'on lui attribuait au sein des sciences cognitives. Il s'agirait au contraire d'utiliser l'intentionnalité comme une ressource pour accroître les expressions de l'intelligence plutôt que de rendre compte de l'intentionnalité, elle-même conçue comme exercice d'une intelligence réduite au calcul. L'intelligence artificielle apparaîtrait, dans ce contexte, comme l'étude scientifique des conditions rendant l'interaction avec les machines plus intelligente et plus complexe, pour que celles-ci deviennent plus manipulables par les êtres humains<sup>7</sup>. L'accent est donc mis sur l'interprétation des situations (et non sur la constitution de règles générales de comportement) et l'étude des types de ressources, représentatives ou autres, convoquées par les humains dans les situations en question<sup>8</sup>. Ainsi l'esprit ne se confond-il plus avec une intériorité qu'on cherche à expliciter, mais il se distribue plutôt dans le complexe instrumental et sémiotique, explicite et matériel, que l'intelligence artificielle contribue à construire. La seconde façon de concevoir le projet de l'intelligence artificielle est plus récente et sans doute moins partagée que la première, qui apparaît désormais comme le paradigme "classique", c'est-à-dire dont les contours théoriques ont été le mieux délimités et qui, par conséquent,

---

<sup>5</sup> Cf. D. Andler, (1990).

<sup>6</sup> Y M Visetti fait remarquer à juste titre que des projets scientifiques comme celui qui consiste à réintégrer la théorie des systèmes dynamiques dans le cadre d'une psychologie computationnelle et dont J. Petitot s'est fait le défenseur demeurent tributaire du paradigme fonctionnaliste. Cf. Y M Visetti, (1994) : 15-48.

<sup>7</sup> Cf. le compte-rendu du livre de L. Suchman, *Plans and Situated Actions - The Problem of Human/Machine Communication* par Y.-M. Visetti, (1989) : 67-96.

<sup>8</sup> Cf. B. Bachimont, (1993) : 213-238.

se prête mieux à la réfutation. Une remarque s'impose toutefois : l'intelligence artificielle première du nom, même si elle est aujourd'hui en perte de vitesse, est liée à la modélisation cognitive. C'est en ce sens qu'elle nous intéresse tout particulièrement ici, puisque c'est sous l'angle de la modélisation cognitive que je vais aborder la question de l'expérience et de l'expérimentation informatique.

D'un point de vue à l'autre, c'est le statut de l'*a priori* qui a évolué : dans le cas de la "science des conditions générales et effectives de l'intelligence", il s'agit de dégager des conditions universelles *a priori* permettant de fonder la notion d'intentionnalité et d'expliquer par ce biais le comportement intelligent. Dans le cas de la "science des artefacts interactifs", c'est l'étude des situations qui révèle, dans l'effectuation, des comportements qui n'étaient pas encadrés, ne serait-ce que virtuellement, par des règles, au sens logico-symbolique et mentaliste. Plus généralement, l'évolution de la notion d'intelligence artificielle semble avoir suivi, avec un "décalage" de cinquante ans, l'évolution historique de la logique qui va de la logique interprétée comme science normative pour les mathématiques à l'informatique théorique comme science du calcul. La logique fut tout d'abord au début du siècle une théorie des fondements des mathématiques ; à partir de son axiomatisation liée au programme de Hilbert dans les années vingt, et même avant les grands théorèmes de limitation interne des formalismes, la logique avait perdu ce rôle normatif de fondement<sup>9</sup> pour devenir une théorie des outils et des champs idéaux à l'œuvre dans les théories mathématiques. En donnant naissance à la théorie du calcul à la fin des années trente puis à la théorie des modèles à la fin des années quarante, la logique a acquis du même coup une "positivité" au sein des théories mathématiques qui l'éloignait un peu plus de la perspective strictement fondationnelle du début du siècle en en faisant un champ d'étude mathématique comme les autres<sup>10</sup>.

Je me limiterai dans la suite au premier sens du terme d'intelligence artificielle, d'une part parce que ses liens avec sa science tutélaire, l'informatique théorique, sont plus clairs et d'autre part parce que ses modélisations ont très directement modifié la notion d'expérience. Cependant, une partie de ce qui sera avancé reste valide, avec les ajustements nécessaires que j'indiquerai, pour l'intelligence artificielle entendue au deuxième sens.

J'en viens donc maintenant à l'expérience.

## 1.2. L'expérience comme rapport à une extériorité

---

<sup>9</sup> Exactement à partir du moment où Hilbert se donnait un champ numérique sans chercher à le réduire à autre chose que lui-même. Cf. D. Hilbert, (1923) : 131-144.

<sup>10</sup> Cf. J. van Heijenoort, (1967).

L'expérience a le statut d'un *rapport à une extériorité* parce que le sujet ne crée pas la réalité et qu'il n'a accès à elle que par le biais de théories qui tentent d'en décrire les conditions de production. L'expérience possède ainsi deux aspects : le premier aspect, celui que l'on rencontre dans l'expression "faire une expérience" met l'accent sur l'absence de maîtrise du sujet à l'égard des conditions de production de la réalité et de ce fait sur la passivité relative du sujet à l'égard de celles-ci ; le deuxième aspect, celui que l'on rencontre dans les expressions "monter une expérience" ou encore "expérimentation" consiste à tenter d'isoler les conditions de production en question et, si possible, à les reproduire. Les deux aspects sont le plus souvent mêlés et je ne tenterai pas de délimiter dans l'abstrait où se situent la passivité et l'activité du sujet dans son rapport à l'expérience.

L'expérience, dans les deux aspects que je viens de dégager, est alors la source d'une *autorité* et c'est dans la mesure où cette autorité est, en droit, *accessible à tous* qu'elle peut intéresser le scientifique. C'est précisément ce qui distingue la modalité scientifique et la modalité mystique de la notion d'expérience : dans ce dernier cas, il s'agit avant tout d'une expérience intime et non-communicable ou au moins dont les conditions d'accès ne sont pas intégralement spécifiées car elles sont irréductiblement subjectives.

Je vais discuter de cinq traits que possède la notion d'expérience dans sa version scientifique, traits que l'on peut diviser en traits globaux et traits locaux :

- D'un point de vue *global*, l'expérience c'est :

1 - la *cohérence* d'un domaine d'événements : il y a une relative homogénéité des événements apparaissant au sein du domaine.

2 - l'*ouverture* indéfinie d'un domaine d'événements : les événements apparaissant au sein du domaine, lui-même borné ou non, sont repérables quoiqu'en nombre indéfini.

- D'un point de vue *local*, les événements se produisant dans le domaine sont susceptibles d'être :

3 - *observables* (selon un cadre d'interprétation fixé à l'avance) : soit par le biais des sens, soit par le biais d'instruments dont les mesures sont interprétées par les sens.

4 - *réitérables* : les événements peuvent se reproduire naturellement ou être reproduits artificiellement (selon le cadre interprétatif choisi à l'avance).

5 - *inattendus* : les événements peuvent ne pas être prévus, que ce soit dans le temps, dans leur forme ou leur intensité (l'aspect inattendu d'un événement peut jouer un rôle important dans la confirmation ou l'infirmité d'une interprétation).

On pourrait critiquer la pertinence de ces cinq traits en faisant remarquer qu'ils ne permettent pas d'avancer dans le débat portant sur la question de savoir quels sont les rôles respectifs de la déduction et de l'induction au sein de la démarche expérimentale. Je dirai volontiers en effet que ces traits sont trop imprécis pour permettre d'entrer dans ce débat — qui ne fait donc pas l'objet des pages qui suivent — mais qu'ils sont en revanche suffisamment précis pour permettre de montrer ce que l'intelligence artificielle a changé dans notre interprétation de la notion d'expérience.

### 1.3. Vocabulaire

Commençons par préciser un certain nombre de points de vocabulaire, sans nous y attarder trop longtemps. Il y a, autour de la notion d'expérience, une nébuleuse d'expressions telles que “sciences expérimentales” ou “méthode expérimentale” dont il faut réussir à préciser le sens parce qu'elles sont au carrefour de deux traditions historiques, l'une française et l'autre anglo-saxonne.

Remarquons tout d'abord que la signification des expressions de “science expérimentale” et de “méthode expérimentale” est en partie déterminée, pour nous francophones, par l'utilisation qu'en fait Claude Bernard [1813 ; 1878] dans ses *Principes de médecine expérimentale* (posthume).

Claude Bernard opère une distinction entre les *sciences d'observation* qui décrivent les phénomènes (généralement hors de portée) et permettent de les prévoir (son exemple favori est l'astronomie) et les *sciences expérimentales* qui non seulement permettent de prévoir les phénomènes mais aussi d'agir sur eux et de les modifier en vue d'une fin qui a pour nous un intérêt vital<sup>11</sup>. Les sciences expérimentales auraient ainsi deux caractéristiques : d'une part, les objets qu'elles étudient auraient un intérêt pragmatique ; d'autre part, ces objets dépendraient, au moins en partie, de l'acte de connaissance lui-même. Cette distinction n'est pas conservée dans l'usage actuel puisqu'aujourd'hui, la notion de science expérimentale ne recouvre pas seulement les sciences qui, comme la médecine ou la biologie, peuvent agir sur et modifier les phénomènes qu'elles étudient mais aussi celles qui ont un rapport à l'extériorité de la nature. Il s'agit là d'un sens plus général, communément répandu dans les pays anglo-saxons et qui dérive de l'œuvre de Francis Bacon [1561 ; 1626]. De ce point de vue, la signification de l'expression “science expérimentale” s'est rapprochée de l'usage anglo-saxon sans se

---

<sup>11</sup> Cf. Claude Bernard, (posthume) : 82.



confondre complètement avec lui (on parlait encore naguère dans les programmes de l'Éducation Nationale des sections "sciences expérimentales" — chimie, biologie — par opposition aux sections dans lesquelles dominaient les mathématiques et la physique).

En revanche, la notion de "méthode expérimentale" semble bien avoir conservé la signification que lui donnait déjà Claude Bernard<sup>12</sup> :

«La méthode expérimentale n'est rien autre chose qu'un ensemble de règles sanctionnées par l'expérience et qui ont pour but de prémunir contre les erreurs qui peuvent résulter du manie ment des faits et des hypothèses dans l'édification de la science».

Pour nous aujourd'hui, "méthode expérimentale" renvoie bien, comme le fait remarquer Claude Bernard, à "un ensemble de règles sanctionnées par l'expérience", dont le rôle reste essentiellement négatif dans la mesure où elles indiquent ce qu'il ne faut pas faire sans préciser ce qu'il faut faire. Par exemple, une de ces règles pourrait être : "on ne peut pas inférer à partir d'un échantillon d'événements trop restreint". Mais les difficultés commencent à surgir quand on tente de définir et de dénombrer ces règles. A ce sujet, personne n'est d'accord pour savoir quelles sont les règles ni même s'il faut admettre qu'il y ait des règles<sup>13</sup>.

#### 1.4. Histoire

On ne peut pas remonter indéfiniment de mot en mot pour éclaircir les significations mais on peut cependant remarquer que le mot d'expérience est un *vieux* mot qui a été employé dans un contexte scientifique bien avant l'avènement des sciences expérimentales ou la mise au jour des principes de la méthode expérimentale. La notion d'expérience utilisée en vue d'une explicitation des phénomènes naturels a cours depuis le XIII<sup>ème</sup> siècle au moins et elle ne commence à prendre son sens actuel qu'avec l'avènement d'une physique théorique, c'est-à-dire au XVII<sup>ème</sup> siècle.

Du XIII<sup>ème</sup> au XVII<sup>ème</sup> siècle, le terme d'expérience revêt un sens essentiellement *hermétique* : recourir aux expériences, c'est, par exemple en alchimie, tenter d'opérer la transmutation des métaux. Le sens du mot "expérience" et le recours à cette notion sont liées, dans ce contexte, à la croyance en la magie de la nature. Chez Paracelse par exemple [1493 ; 1541], la nature est magique et inversement, la magie est une procédure naturelle d'explicitation. Qui peut savoir à l'avance ce que peut produire l'influence des astres et des éléments sur la nature physique ? Seules les

---

<sup>12</sup> Claude Bernard, (posthume) : 78.

<sup>13</sup> Cf., pour un point de vue provocateur, R. Thom, (1986) : 7-20.

*expériences* peuvent le manifester. Comme le fait remarquer Koyré<sup>14</sup>, pour Paracelse, le boulanger est un alchimiste quand il transforme la farine et le levain en pain, le pharmacien est un alchimiste quand il prépare des drogues en extrayant “l’esprit” des différentes matières ; le corps lui-même est un alchimiste qui sépare dans les éléments ce qui peut servir de nourriture. Bref, dans le contexte de l’alchimie, faire des expériences consiste à rendre plus rapide un processus naturel, la nature ultime étant la transmutation elle-même, telle qu’elle est symbolisée par l’incarnation du Christ. On est loin d’un recours à l’expérience au sens où nous l’entendons ...

Remarquons cependant qu’il y a au moins un point commun entre notre attitude scientifique “moderne” et l’attitude pré-moderne des alchimistes de la Renaissance : comme nous l’avons souligné en commençant, l’expérience est une *source d’autorité* puisqu’on se réfère à elle comme à une “pierre de touche” (autre expression philosophique empruntée aux investigations touchant la nature des métaux) permettant de vérifier la valeur de nos connaissances. Il y a donc chez eux comme chez nous, une *foi* en l’expérience : cette foi n’est donc pas l’apanage de la rationalité moderne.

On conçoit donc facilement d’où provient la difficulté d’une définition de l’expérience : si l’apparition de la notion avait coïncidé avec l’avènement de la science moderne, on aurait pu réussir à dégager un sens univoque et en donner peut-être même une définition. Mais ce n’est pas le cas, comme le montre l’exemple de l’alchimie. Il faut donc essayer de voir comment la signification du recours à l’expérience s’est transformée au cours du temps et comment on est passé du sens ancien au sens moderne. On tentera ensuite de voir si le sens moderne du mot s’est lui-même transformé avec l’avènement de nouvelles branches du savoir scientifique comme l’intelligence artificielle.

## 2. APERÇU HISTORIQUE. LES TROIS POINTS DE METHODE TOUCHANT LE RAPPORT A L’EXPERIENCE

Pour souligner les différences entre le point de vue ancien et le point de vue moderne touchant la notion d’expérience sans toutefois entrer dans le détail du développement des sciences pendant une période aussi longue, j’ai recours à trois traits caractéristiques touchant les méthodes utilisées dans le recours à l’expérience, suivant en cela R. Blanché<sup>15</sup>. Il faudra essayer de dégager les interprétations successives que l’on a donné :

1. Du raisonnement hypothético-déductif
2. De la mathématisation

---

<sup>14</sup> A. Koyré, (1971) : 100 sq.

<sup>15</sup> R. Blanché (1969).

### 3. De l'instrumentation

#### 2.1. Le raisonnement hypothético-déductif dans l'antiquité et à l'époque moderne

La logique du raisonnement hypothético-déductif a été dégagée dès l'antiquité grecque et c'est aussi dès cette époque que ce raisonnement a été appliqué aux investigations touchant la nature physique. Mais ce sont le statut et la valeur accordée à la notion d'hypothèse qui ont changé par rapport à l'époque moderne. Dans l'antiquité, la notion d'hypothèse emprunte son sens à la géométrie, considérée comme la science décrivant idéalement les rapports entre les objets de la réalité, idéale ou physique. La réalité physique est, dans une certaine mesure, susceptible d'être décrite en termes géométriques, du moins est-ce l'une des tâches que s'assigne la science grecque.

Le présupposé épistémologique fondamental qui se dégage de cette attitude scientifique, c'est qu'*on ne croit pas* en la vérité des hypothèses mathématiques qui sont émises : elles sont seulement un moyen de "sauver les phénomènes", c'est-à-dire de rendre compte des apparences issues de la perception. Ce que j'ai appelé en commençant la "foi en l'expérience" est donc ambivalente : d'une part, on accorde foi en la perception des sens et non dans les hypothèses mathématiques que l'on propose mais d'autre part, on considère le sensible perçu comme irrémédiablement irrationnel de par le caractère approximatif de l'analyse géométrique du mouvement que l'on peut effectuer sur lui. On peut donc dire que, pour les anciens dont le canon scientifique est l'axiomatisation de la géométrie interprétée comme "voie de salut" de la rationalisation du sensible, les sciences ayant rapport à l'expérience telles que l'astronomie sont des *axiomatiques imaginaires* puisqu'elles sont considérées comme de simples hypothèses dont on ne peut pas connaître la valeur de vérité, étant donné l'aspect irrationnel du statut de la perception du mouvement dans l'univers physique.

Le raisonnement hypothético-déductif moderne change profondément de nature puisque les hypothèses sont réintégrées dans le cadre logique du vrai et du faux tout en étant dissociées d'une approche d'emblée axiomatique. Le véritable initiateur de ce nouveau mode de pensée est Galilée [1564 ; 1642]. C'est lui qui tente de reconstruire le monde physique à partir d'hypothèses considérées comme vraies en imaginant des expériences visant à corroborer les hypothèses qu'il formule touchant la nature physique. Remarquons que les expériences de Galilée sont des *expériences de pensée* et que la confirmation issue de l'expérience peut, dans certains cas, attendre des années ou même ne

pas avoir lieu du tout (Galilée se vantait auprès de certains de ses correspondants de ne pas avoir besoin du contrôle de l'expérience). On a beaucoup discuté la question de savoir si Galilée faisait réellement ou non les expériences qu'il décrivait ou s'il se contentait de faire des expériences de pensée<sup>16</sup>. Sans vouloir entrer dans le débat, je me contenterai seulement de remarquer deux points.

D'une part, la composante imaginaire présente dans la notion d'expérience dès l'origine n'a pas été abolie par l'avènement de la science moderne, puisqu'il y a un sens à invoquer dans ce nouveau cadre la notion d'expérience de pensée. D'autre part, contrairement aux axiomatiques imaginaires des Anciens, l'imaginaire présent dans la notion d'expérience de pensée vise bien à émettre une hypothèse portant sur la nature physique *en la réintégrant dans la sphère du vrai et du faux*<sup>17</sup>.

On peut dire qu'à cette époque, le statut à accorder à l'expérience en physique pose deux types de questions. D'une part, se pose la question du statut *axiomatique* à accorder à la physique : peut-on, en physique, constituer une science sur le modèle de l'axiomatique euclidienne, c'est-à-dire une science démonstrative qui dérive ses propositions d'un petit nombre de propositions indémontrables que l'on aurait réussi à isoler ? D'autre part, la physique est-elle une science nécessaire ou bien ses méthodes sont-elles essentiellement probabilistes ? Autrement dit, l'obligation qu'il y a de se référer à l'expérience en physique limite-t-elle la portée des principes de la physique à des domaines essentiellement *locaux* ?

Ces deux types de question sont évidemment liés et traversent toute la philosophie de l'âge classique. Les réponses sont variables et si ce que l'on a pris l'habitude d'appeler le "rationalisme classique" a penché dès l'origine pour considérer la physique comme une science nécessaire à vocation universelle (Descartes [1596 ; 1650], Newton [1642 ; 1727] et Kant [1724 ; 1804] font partie de ce courant), il y a eu cependant, dès le début, des positions contraires, comme celle de Pascal qui considère que la physique n'a pas besoin d'un fondement axiomatique *a priori* dans la mesure où il

---

<sup>16</sup> Cf. Pour ce débat d'érudition A. Koyré (1973) : 213-223 et R. Dugas (1954) : 82-83.

<sup>17</sup> Les astronomes arabes de l'époque médiévale avaient, par souci "réaliste", tenté de construire une astronomie qui ne soit pas imaginaire. Mais ils s'étaient reposés pour ce faire sur la doctrine aristotélicienne. Cf. P. Duhem, (1908) : 31 : «Au contraire, avec Thâbit ibn Kourrah, avec Ibn-al-Haitam, les astronomes arabes ont voulu que les hypothèses qu'ils formulaient correspondissent à des mouvements véritables de corps solides ou fluides réellement existants ; dès lors, ils ont rendu ces hypothèses justiciables des lois posées par la Physique. Or la Physique professée par la plupart des philosophes de l'Islam était la Physique péripatéticienne [...]».

s'agit *précisément* d'une science expérimentale : pour lui, les différents types d'expériences permettent de manipuler des *rappports entre les objets* selon des règles élaborées par les êtres vivants que nous sommes, une fois que l'on a posé des définitions linguistiques conventionnelles touchant le statut de ces objets<sup>18</sup>. D'autres auteurs, comme Hume [1711 ; 1776], ont eux aussi tenté de s'opposer à la "foi rationaliste" en l'aspect axiomatisable de la physique.

Quels que soient les tenants et les aboutissants de ce débat interne à l'âge classique, on doit retenir que le raisonnement hypothético-déductif a subi un changement d'interprétation majeur au cours du XVII<sup>ème</sup> siècle. En effet, c'est bien le projet de déterminer *la vérité ou la fausseté* d'une question posée à la nature par le biais d'un montage expérimental qui retient désormais l'attention expérimentale des physiciens et l'attention logique des philosophes.

## 2.2. La mathématisation pré-moderne et moderne

La fascination pour la mathématisation des phénomènes est sans doute l'un des traits les plus constants de la tradition occidentale depuis Pythagore [VI<sup>ème</sup> siècle avant l'ère chrétienne] et il n'y a rien de "moderne" dans cette attitude qui consiste essentiellement à trouver une forme mathématique susceptible de décrire les phénomènes tels qu'ils sont *perçus*. C'est une attitude qui a été fort répandue au moyen-âge dans la tradition de la cabale et jusqu'à Kepler [1571-1630].

Dans les textes de Kepler par exemple, on trouve encore des justifications mathématiques à la description du monde tel qu'il apparaît. Kepler reprend par exemple la théorie géométrique des polyèdres réguliers pour décrire la distance des planètes entre elles. Pourquoi se sert-il de la théorie des polyèdres réguliers ? On peut invoquer deux raisons à ce fait. C'est Platon qui le premier s'est servi de cette théorie pour rendre compte, dans le *Timée*, des formes présentes dans la nature. Plus tard, c'est pour rendre raison de la théorie des polyèdres réguliers qu'Euclide a construit son axiomatique de la géométrie. En se servant de la théorie des polyèdres réguliers, Kepler, même s'il ne cherche pas expressément le patronage illustre de l'autorité des anciens, adopte une attitude d'esprit visant l'axiomatisation *a priori* de l'astronomie. Un de ses textes est, à cet égard, particulièrement significatif ; à l'orbe de la Terre,

«circonscris un dodécaèdre : la sphère qui l'entoure est celle de Mars. À l'orbe de Mars, circonscris un tétraèdre : la sphère qui l'entoure est Jupiter. À l'orbe de Jupiter, circonscris un cube : la sphère qui l'entoure est Saturne. Place maintenant dans l'orbe de la Terre un icosaèdre : la sphère qui lui est

---

<sup>18</sup> Cf. C. Chevalley, (1995) : 91 sq.

inscrite est Vénus. Place dans l'orbe de Vénus un octaèdre : la sphère qui lui est inscrite est Mercure. Telle est la raison du nombre des planètes.<sup>19</sup>»

On voit donc ici que la mathématique sert essentiellement à justifier une approche axiomatique de l'astronomie dont le but est de plier l'expérience à une interprétation déjà toute faite : dans cet exemple, l'aspect contingent du nombre des planètes trouve ainsi une pseudo-justification dans une application arbitraire d'un certain nombre de figures géométriques aux distances entre les planètes. Il s'agit donc de donner une *apparence* géométrique aux formes *telles qu'elles sont perçues*. On peut parler dans ce cas d'une *application* des mathématiques à la réalité mais pas d'une mathématisation objective de la réalité, qui suppose, comme on va le voir maintenant, un autre rapport entre mathématiques et réalité.

La mathématisation moderne est toute différente et ne consiste pas en une application des mathématiques à la réalité perçue. La première caractéristique de cette mathématisation est la *méfiance* à l'égard de la perception par les sens. Chez Galilée déjà, les invariants physiques ne sont plus accessibles directement à la perception dans la mesure où ils n'ont de signification que dans des rapports mathématisés.

Une fois que l'on admet que l'invariant physique n'a d'existence qu'en tant qu'objet mathématique idéalisé et non pas en tant qu'objet directement perçu, on comprend pourquoi le discours mathématique n'est pas extérieur au réel et ne "porte" pas "sur" lui : le domaine de l'objectivité peut être défini comme celui d'une construction mathématique de la réalité et non pas comme une simple description mathématique. La distinction fondamentale se situe entre d'une part une application des mathématiques sans générativité, application qui ne permet pas de prévision puisqu'il s'agit seulement d'une description du sensible perçu interprétée comme matérialisation imparfaite de formes géométriques et d'autre part, une construction mathématique du réel permettant une générativité qui permet elle-même une prévision. C'est dans ce cadre mathématique qu'il devient possible de prévoir les résultats des expériences et d'établir s'il y a ou non écart entre le modèle mathématique *a priori* qui sert de norme et l'expérience telle qu'elle est réalisée. Sans théorie mathématique servant de norme, l'existence de cet écart ainsi que sa mesure n'ont évidemment aucun sens.

Le vocabulaire de l'*application* (des mathématiques) à (une réalité), est trompeur parce qu'il laisse croire qu'il y aurait une réalité brute qui n'aurait pas besoin d'être "travaillée", ne serait-ce que par les sens, pour accéder au statut de réalité. Or la construction mathématique du réel consiste essentiellement en l'étude des relations au sein d'un espace

---

<sup>19</sup> Cité par A. Koyré, (1961) : 146.

abstrait que l'on a construit pour la circonstance : par exemple, pour déterminer le mouvement d'un point matériel dans l'espace, il faut connaître sa position et sa vitesse, ce qui exige de faire intervenir un espace à six dimensions, qui n'a donc plus qu'un rapport d'analogie avec l'espace physique tel qu'il est perçu. Aussi la possibilité d'une mesure est-elle conditionnée par la construction de cet espace abstrait. C'est pour rendre compte de cette différence entre espace mathématique et espace perçu que l'on a introduit dès le XVII<sup>ème</sup> siècle la distinction entre les qualités premières susceptibles d'être mathématisées (étendue, mouvement) et les qualités secondes qui ne le sont pas (couleur, odeur, forme, etc.). Il faut donc dépouiller la nature de tout le sensible contingent<sup>20</sup> (par exemple, dans l'expérience de la chute des corps telle qu'elle est décrite par Galilée, la résistance de l'air), en faisant intervenir la différence entre les qualités premières et les qualités secondes<sup>21</sup>.

### 2.3. Outil et instrument

Pour comprendre la nature du rôle des instruments dans la notion d'expérience, il est commode d'opérer une distinction entre l'*outil* qui n'est pas conçu selon un plan mathématique et l'*instrument* qui l'est.

L'usage d'outils dans le rapport à la nature semble aussi vieux que l'humanité et sans doute plus vieux encore<sup>22</sup>, ce qui laisse supposer l'existence d'une prédisposition biologique à la construction d'outils. Le rapport à la nature et partant, sa connaissance, semble donc d'emblée supposer, pour l'espèce humaine, la médiation de l'outil.

On remarque cependant que l'antiquité grecque et romaine concevait l'outil comme appartenant au registre de l'action pratique *par opposition* à la connaissance théorique, ce qui avait pour conséquence d'empêcher qu'il puisse y avoir, par le biais d'outils, une connaissance théorique portant sur la nature. Cette distinction nette entre l'action pratique et la connaissance théorique provient sans doute du fait que l'action était interprétée comme se

---

<sup>20</sup> Cette contingence pose néanmoins de redoutables problèmes épistémologiques puisque la décomposition en qualités premières — seules scientifiques — et secondes prive *par définition* l'observateur d'une analyse scientifique de la causalité des phénomènes associés aux qualités secondes.

<sup>21</sup> Le modèle épistémologique de cette différence se trouve chez Descartes dans son exemple de la décomposition d'un morceau de cire, bien que les termes de "qualité première" et "qualité seconde" soient postérieurs. Cf. Descartes, (1641) : 2<sup>ème</sup> méditation.

<sup>22</sup> Si l'on en croit Y. Coppens en effet, l'usage d'outils est moins une caractéristique propre aux humains qu'aux hominidés. Cf. Y. Coppens, (1983) : 222.

déroulant dans le temps alors que la connaissance théorique était interprétée comme tentant de rendre compte de la nature de l'espace (le modèle de cette connaissance est, on le sait, la géométrie). Aussi l'usage d'outils dans le rapport à la nature n'était-il pas conçu selon un plan mathématique ni comme la matérialisation d'une règle mathématique au sein de la nature, comme cela se généralisera à partir du XVII<sup>ème</sup> siècle<sup>23</sup>.

L'idée d'instrument est donc bien antérieure à la mathématisation du réel telle qu'elle a eu cours au XVII<sup>ème</sup> siècle, même si elle n'était pas l'attitude dominante. En quoi consiste la généralisation de la fabrication et de l'utilisation des instruments ?

Il y a instrument scientifique quand on construit un outil en anticipant ses effets par le biais d'une représentation de type mathématique. L'outil est alors la réalisation particulière et reproductible d'un type idéal représenté par l'instrument<sup>24</sup>. C'est cette nouvelle attitude à l'égard du couple outil / instrument qui apparaît dans les milieux scientifiques au cours du XVII<sup>ème</sup> siècle : on voit Descartes, Pascal ou Huygens demander à des artisans de leur construire des instruments selon des plans mathématiques et exiger d'eux une précision de plus en plus grande dans leur fabrication. Pascal par exemple réalise ainsi plus

---

<sup>23</sup> C'est le cas chez Descartes (cf. (1637), II) et chez Pascal (cf. (1670) n° 486). Un contre-exemple majeur peut cependant être évoqué à l'appui de la thèse de l'existence d'instruments dans les sciences rationnelles avant le XVII<sup>ème</sup> siècle. Dans l'antiquité, on connaît en effet le cas d'Archimède [287; 212] qui construisit bel et bien des instruments en les pensant comme des matérialisations de règles mathématiques. La tradition a surtout conservé l'aspect merveilleux des machines de guerre construites par Archimède lors du siège de sa ville, Syracuse, par l'armée romaine. Mais il y a bien plus dans l'œuvre d'Archimède, comme le montre son traité *Méthode concernant les théorèmes de mécanique*, redécouvert en 1906 : pour la première fois, Archimède y conçoit une nouvelle manière de faire des mathématiques en montrant comment certains théorèmes «sont mis en lumière grâce à la mécanique». C'est par une expérience mentale de *mise en mouvement* de certaines figures les unes par rapport aux autres qu'il est possible d'aboutir à la construction d'objets mathématiques nouveaux. C'est cette nouvelle attitude touchant les objets mathématiques, conçus non plus comme des objets à contempler — comme chez Platon ou Euclide — mais comme résultant d'une *construction temporelle réglée*, qui va se généraliser au XVII<sup>ème</sup> siècle.

Cf. D. R. Lachterman, (1989).

<sup>24</sup> Voici par exemple ce que dit Pascal de sa machine arithmétique dans un vocabulaire différent du nôtre, où la notion de modèle joue le rôle que nous attribuons à l'instrument, (1645) : 188 : «Les lumières de la géométrie, de la physique et de la mécanique m'en fournirent le dessein et m'assurèrent que l'usage en serait infallible si quelque ouvrier pouvait former l'instrument dont j'avais imaginé le modèle».



d'une cinquantaine de machines à calculer avant de parvenir au modèle définitif de sa machine<sup>25</sup>. C'est encore Galilée qui semble être l'initiateur de cette nouvelle attitude et l'on peut dire de ce point de vue que sa lunette astronomique est le premier instrument moderne. L'usage empirique de la lunette, c'est de voir plus gros et donc plus loin (c'est cette utilisation qui intéressait le Doge de Venise pour assurer la surveillance des côtes de la République quand Galilée lui présenta sa lunette). Mais il s'agit là d'une simple *application* empirique de l'instrument, conçu comme un prolongement du corps qui le transforme en outil plus efficace mais qui ne transforme pas son rapport au monde. Or l'instrumentation, de même que la mathématisation, n'est pas une simple "application" de recettes mathématiques à une réalité qui lui préexisterait.

L'usage proprement astronomique de la lunette qui la transforme en instrument est en effet différent : cet usage permet certes de voir plus gros mais il permet surtout de déterminer le rapport *mathématique* entre ce que l'on voit à l'œil nu et ce que l'on est susceptible de voir idéalement grâce à l'instrument, c'est-à-dire de constituer *en retour* l'œil humain en un instrument ayant certaines propriétés mathématiques déterminables. On peut dire, en se fiant à cet exemple, qu'il y a une généralisation de la pratique instrumentale à partir du moment où l'on a pu envisager les performances du corps humain lui-même à partir d'une norme mathématique. Le corps devient alors un instrument particulier qui obéit à certaines lois mathématiques : par exemple, dans le cas de cette partie du corps qu'est l'œil, les lois de l'optique rendent compte de la possibilité de cette fonction particulière qu'est la vision<sup>26</sup>. On voit que l'outil n'est plus alors pensé comme un simple prolongement du corps ayant but d'en accroître les potentialités mais que *le corps entre lui-même dans un schéma abstrait dont il est l'un des produits possibles*. L'idée cartésienne d'un corps-machine découle de cette attitude nouvelle à l'égard de l'instrument.

Les trois caractéristiques que nous venons d'analyser ont permis de situer au XVII<sup>e</sup> siècle un changement majeur dans la signification accordée à la notion d'expérience. Les conséquences de cette transformation se font sentir encore aujourd'hui : peuvent en témoigner des textes aussi divers épistémologiquement que *La crise des sciences européennes et la philosophie transcendantale* de Husserl [1859 ;

---

<sup>25</sup> Cf. Pascal, (1645) : 191.

<sup>26</sup> Que les premiers actes de chirurgie sur le corps humain se soient déroulés à la même époque et dans les mêmes lieux que ce que l'on a appelé la "rupture galiléenne" n'est sans doute pas un hasard. Cf. G. Canguilhem, (1965).

1938] qui décrit la “rupture galiléenne” du XVII<sup>ème</sup> siècle ou encore *Les fondements logiques des probabilités* de Carnap [1891 ; 1971] qui construit une logique des probabilités indispensable, d’un point de vue logique, à l’aspect conjectural des sciences expérimentales modernes.

Une constatation frappante touchant les sciences d’aujourd’hui est l’utilisation massive qu’elles font des ordinateurs en tant qu’outils de calcul. Mais limiter leur usage à cette fonction tend à faire oublier la science mathématique qui fonde la constitution de cet outil, à savoir l’informatique théorique. Il y a en effet bien plus dans l’utilisation de l’informatique que son usage empirique — tangible dans l’ordinateur en tant qu’outil de calcul — usage qui n’est pas essentiellement différent de celui que le Doge de Venise réservait à la lunette de Galilée : un instrument prolongeant le corps et décuplant ses potentialités. Si l’informatique joue bien le rôle d’un instrument dans la science d’aujourd’hui, c’est le rapport du scientifique au monde, c’est-à-dire la constitution de ce que l’on entend par réalité, qui doit s’en trouver changée. Ce changement dans la réalité doit affecter à la fois la constitution de l’objectivité et la nature de l’empiricité. C’est précisément sur ce double changement qu’il faut s’interroger quand on tente d’analyser les rapports entre la notion d’expérience et celle d’intelligence artificielle.

De ce point de vue, Allen Newell et Herbert A. Simon ont tracé la voie à suivre dans leur célèbre article de 1976 “Computer Science as Empirical Enquiry : Symbols and Search” qui développe deux idées capitales touchant l’intelligence artificielle : d’une part le support de l’informatique relève du symbolique en tant qu’il est manipulation d’une écriture ; d’autre part, le domaine du symbolique permet d’aborder les différentes sciences humaines, de la psychologie à l’économie, d’un point de vue fonctionnel, c’est-à-dire en se mettant à distance des contenus réels de ces disciplines. Voici ce qu’ils écrivent<sup>27</sup> :

«Nous construisons des ordinateurs et des programmes pour des raisons multiples. Nous les construisons en vue de servir la société et en tant qu’outils permettant de mener à bien les tâches économiques qu’elle s’est fixée. Mais en tant que scientifiques purs, nous construisons des ordinateurs et des programmes *en vue de découvrir des phénomènes nouveaux*<sup>28</sup> et d’analyser des phénomènes que nous connaissons déjà par ailleurs. La société s’égare souvent sur ce point quand elle croit que les ordinateurs et les programmes ne sont construits que pour l’usage économique que l’on peut en faire (ou comme des intermédiaires dans une chaîne de

<sup>27</sup> Newell et Simon, (1976) : 105-132.

<sup>28</sup> C’est moi qui souligne.

développement conduisant à cet usage). Elle doit comprendre que les phénomènes qui entourent les ordinateurs sont profonds et obscurs, qu'ils requièrent beaucoup d'expérimentation pour qu'on établisse leur nature. Elle doit comprendre que, comme dans toute science, les bénéfices que l'on tire de cette expérimentation et de cette compréhension deviennent payants pour ce qui est de l'acquisition permanente de nouvelles techniques et que ce sont ces techniques qui créeront les instruments permettant à la société de réaliser ses buts.»

Je vais en venir maintenant à ce que l'intelligence artificielle semble avoir fait changer dans la notion d'expérience.

### 3. LES NOTIONS D'EXPERIENCE ET DE MODELISATION TELLES QU'ELLES ONT ETE TRANSFORMEES PAR L'INFORMATIQUE ET ILLUSTRÉES PAR L'INTELLIGENCE ARTIFICIELLE

Si l'on en revient à notre première description de la notion d'expérience caractérisée comme ayant le statut d'un rapport à une extériorité parce que le sujet ne maîtrise pas les conditions de production de la réalité, on s'aperçoit que l'intelligence artificielle a profondément transformé cet aspect fondamental de la notion. Nous allons tenter de justifier l'idée que *l'extériorité dont il est question en intelligence artificielle n'est plus seulement l'extériorité de la nature mais qu'il est légitime de prendre en considération une extériorité interne à la pensée*. C'est donc autour de la notion d'expérience de pensée que se joue le nouveau sens de la notion d'expérience dans le contexte de l'intelligence artificielle. Remarquons toutefois que cette nouvelle façon d'aborder la notion d'expérience de pensée serait insuffisante pour caractériser l'intelligence artificielle au second sens explicite dans l'introduction : une problématique de la trace et de l'écriture conçue comme l'extériorité matérielle de l'esprit devrait être également évoquée<sup>29</sup>. On pourrait objecter également que la notion d'expérience de pensée existait déjà en mathématiques depuis la nuit des temps. C'est vrai. Mais en mathématiques, ce n'est pas l'expérience de *pensée* qui focalise l'attention mais les *objets* pensés, le mode de pensée des objets en question n'étant pas thématiquement en propre comme c'est le cas en intelligence artificielle.

Le problème du mode d'accès aux objets mathématiques par la pensée définit sans doute le champ philosophique, depuis l'origine platonicienne de la philosophie. De ce point de vue, la prévention du public philosophique à l'égard de l'intelligence artificielle se nourrit également du sentiment *très réel* que cette dernière empiète directement sur un terrain qui est

---

<sup>29</sup> B. Stiegler défend ce point de vue dans (1994), (1996).

traditionnellement réservé à la philosophie et que celle-ci n'interprète pas comme directement technique. Au risque de scandaliser les puristes, je prends le risque d'interpréter les rapports — fluctuants — entre les trois disciplines, mathématiques, philosophie et intelligence artificielle, de la manière suivante. Les mathématiques pensent des *objets* idéaux par le biais d'un entraînement intellectuel qui a deux caractéristiques : d'une part, il est long et ardu d'un point de vue psychologique mais ne vise pas à penser la pensée ; d'autre part, il est éminemment technique et formel — comme en témoigne la tradition vénérable des manuels de mathématiques d'Euclide à Bourbaki. La philosophie considère que le fait même d'avoir une pensée ouvre un champ d'investigation tout à fait spécifique qui n'est pas axé vers la seule détermination d'objets : une pensée, au sens propre, est déjà, en acte, un *hybride* qui manifeste en lui dans l'après-coup la coexistence d'un sujet et d'un objet sous la forme d'un questionnement. L'investigation philosophique du champ de la pensée n'est donc pas de nature mathématique et, partant, n'est pas technique. L'intelligence artificielle se situe au confluent de ces deux disciplines : d'une part, elle s'oppose aux mathématiques dans la mesure où elle cherche à penser la pensée et ceci sur un mode instrumental mais elle hérite des mathématiques l'idée qu'il y a bien un problème technique à résoudre dans la façon dont la pensée peut appréhender des objets, idéaux ou matériels (y compris elle-même) ; d'autre part, elle s'oppose à la philosophie en cherchant à mener une enquête technique sur la nature de la pensée mais elle hérite de la philosophie l'idée qu'une investigation touchant la nature de la pensée est possible par le biais de la pensée, c'est-à-dire qu'une réflexion sur les processus de la pensée est possible.

J'avais mentionné en commençant cinq traits touchant la notion d'expérience. Pour réussir à préciser comment ces cinq traits ont été transformés dans ce nouveau contexte scientifique, je vais tout d'abord montrer comment les trois points de méthode touchant le rapport à l'expérience ont eux-mêmes évolués.

### **3.1. Les trois points de méthode touchant le rapport à l'expérience dans le contexte de l'intelligence artificielle**

#### *3.1.1. Le raisonnement hypothético-déductif : la notion de modèle*

La question de la place du raisonnement hypothético-déductif — qui était de savoir quel statut accorder aux hypothèses des sciences de la nature par rapport au vrai et au faux — prend des formes nouvelles dans le contexte de l'intelligence artificielle. Le débat, comme je viens de l'indiquer, tourne autour de notion de *modèle*.

Un modèle en intelligence artificielle n'est ni une axiomatique imaginaire à la manière des anciens, ni seulement une reconstruction abstraite à la manière des modernes, mais une *interprétation*

*méthodique* informatiquement régulée. Pour saisir le sens de cette expression, prenons comme exemple la façon dont B. Victorri décrit ses modèles connexionnistes de la mémoire<sup>30</sup> :

«[Ils] permettent de simuler, de manière certes très simpliste, des comportements du type de ceux que l'on observe en psychologie expérimentale, à l'aide de systèmes dont l'organisation paraît proche, à un certain niveau d'abstraction en tout cas, des systèmes physico-chimiques que nous révèle la neuro-physiologie. Ainsi ces modèles semblent à terme être en mesure de combler le fossé entre fonctionnement du cerveau et fonctionnement de l'esprit.»

Le modèle semble donc jouer un rôle *articulatoire* qui se décompose en deux.

D'une part, un modèle n'essaye pas tant de constituer un champ d'objets que de relier au sein d'un espace *ad hoc* des données recueillies par d'autres disciplines au moyen de méthodes qui leur sont propres et sur lesquelles le modélisateur n'a pas — ou a peu — de prise directe. Dans le cas décrit à l'instant, il s'agit de rendre compte de données qui ont été élaborées par une autre discipline, la psychologie expérimentale et qu'il faut accepter, au moins au départ, comme telles, à moins d'avoir au préalable recueilli soi-même des données de psychologie expérimentale. L'extériorité se présente donc moins comme un rapport direct à la nature que comme ce qui provient d'autres formes de savoir auxquelles on est obligé de faire une relative confiance. Ce n'est pas le cas en physique par exemple où les données sont élaborées à partir d'instruments conçus par les physiciens eux-mêmes : l'homogénéité de la discipline apparaît donc plus grande.

D'autre part, en tant qu'il tente de coordonner des points de vue sur le même objet, le modèle induit immédiatement un point de vue épistémologique : la clarification des méthodes permettant de déterminer l'objet interprété par le modèle. Dans l'exemple choisi, cela revient à examiner sous un jour particulier la distinction cerveau / esprit, sans chercher à réduire l'un des termes à l'autre.

Ces deux traits confèrent au rôle articulatoire de la notion de modèle un aspect problématique : en effet, si le but visé par le modèle est, comme l'indiquait B. Victorri, de "combler un fossé" entre deux points de vue sur le même objet, alors ce but paraît au premier abord irréalisable puisque les deux points de vue ont été distingués au préalable et, pour ainsi dire, par définition. Il semble donc quelque peu paradoxal de vouloir à la fois conserver et dépasser cette opposition de

---

<sup>30</sup> B. Victorri, (1995) : 339.

points de vue. En fait, le paradoxe n'en est pas un : la différence de point de vue entre le cerveau et l'esprit repose sur la différence entre les disciplines scientifiques qui la sous-tendent, neuro-physiologie d'une part, psychologie expérimentale de l'autre. Or les frontières entre les disciplines ne sont ni étanches, ni inexistantes et elles bougent avec le temps : le rôle du modèle est précisément d'opérer une double évaluation, d'une part vis-à-vis des *points de passage* entre les disciplines en tentant de dresser la liste des contraintes structurelles qui pèsent sur la possibilité du passage en question et d'autre part vis-à-vis des *résistances* qui se font jour entre les disciplines en tentant de mesurer ce que peut vouloir dire l'idée d'unité de la science, c'est-à-dire l'unité de la démarche propre au sujet connaissant.

La notion de modèle a donc une double fonction : elle a une fonction *scientifique* car elle permet de constituer le terrain (mathématisé) au sein duquel la rencontre de plusieurs disciplines devient pertinente mais elle a aussi une fonction *réflexive* de nature épistémologique dans la mesure où elle permet d'étudier la fonction de connaissance elle-même.

On voit quel intérêt la notion de modèle — et la double fonction qui la caractérise — peut avoir en sciences cognitives puisque c'est par son biais qu'il devient légitime de se demander en quoi consiste un processus physique intelligent. Le but des modèles consiste précisément à *se faire une idée* de ce qu'est un processus physique intelligent<sup>31</sup> sans qu'il y ait à proprement parler reconstruction abstraite du processus en question — ce qui serait le cas si une véritable physique de l'esprit était constituée. Cette reconstruction impliquerait en effet la possibilité d'une générativité du modèle liée à sa forme mathématique et à partir d'elle la possibilité de prévoir les étapes futures du processus physique décrit. Or le modèle n'a pas habituellement les moyens de parvenir à un tel état.

La difficulté qui caractérise la notion de modèle quand on essaye de préciser sa place par rapport à la question du vrai et du faux est donc bien contenue toute entière dans l'interprétation qu'il faut donner au mot de "simulation". La simulation se situe à *distance* de l'objet à connaître<sup>32</sup> ; elle n'est ni la reconstruction nécessaire et causale de

---

<sup>31</sup> B. Victorri le faisait remarquer à la fin de (1995) : 339 : «L'intérêt de ces modèles] ne réside pas dans leur capacité à rendre compte de performances quantitatives mesurables. Il est avant tout de tester des idées, d'ouvrir des pistes de réflexion qui peuvent être utiles aux spécialistes des disciplines concernées.»

<sup>32</sup> Un modèle apparaît alors comme un intermédiaire entre une maquette miniature et une observation grandeur nature. Par exemple, on peut faire passer des test de

l'objet ni sa description intuitive. Elle entretient donc avec la question du vrai et du faux un rapport spécifique que l'on peut caractériser comme étant un "comme si"<sup>33</sup> : par le biais d'un modèle, tout se passe comme si un processus cognitif était mathématiquement décrit depuis sa base physique jusqu'à son rôle représentationnel (par exemple, dans la reconnaissance de formes ou la constitution d'une sémantique).

On peut certes envisager le rôle de la notion de modèle en intelligence artificielle de façon plus ambitieuse et accorder un véritable poids ontologique à ses modèles, en considérant que ceux-ci visent en fait à devenir *identiques* à la *réalité* des processus de la pensée : ils ne décriraient pas seulement les processus de pensée mais seraient eux-mêmes les processus en question<sup>34</sup>. Deux questions doivent être considérées dans cette optique, la première de l'ordre du fait et la seconde de l'ordre du souhait.

De l'ordre du fait. A-t-on déjà réussi à constituer un modèle informatique dont on ait la preuve tangible qu'il soit adéquat au processus de pensée qu'il simule ? Force est de constater que la réponse est *non*, quelles que soient les rodomontades que l'on a pu entendre répéter depuis quarante ans que

---

résistance à une maquette d'avion dans une soufflerie mais on peut aussi faire passer des tests à un avion virtuel en simulant des conditions de vol.

<sup>33</sup> J'emprunte à Y.-M. Visetti le sens de cette locution qu'il utilise dans le cadre de ses réflexions sur la nature de la modélisation en sciences cognitives en le décalquant de l'usage kantien.

<sup>34</sup> C'est la thèse que Searle a baptisé "intelligence artificielle forte" et qu'il a longuement critiquée, en particulier grâce à l'argument de la chambre chinoise. Cf. Searle (1980). L'"hypothèse" de Fodor selon laquelle il y aurait un langage de la pensée me paraît relever de la même attitude. Fodor résume son hypothèse de la manière suivante : 1°. Les opérations mentales sont toutes des systèmes de représentations assorties de moyens de calcul 2°. Ces systèmes ne peuvent pas être des langues naturelles 3°. Il doit donc y avoir un système primitif et inné (J. Fodor, (1975) : 99). En faisant de la pensée un langage doté de "bonnes" propriétés (langage de la pensée *nécessaire* par opposition à l'aspect conventionnel des langues, langage de la pensée suffisamment *riche* pour rendre compte de tous les prédicats des langues, apprentissage défini comme rendant possible la maîtrise de l'extension des prédicats des langues), on résout le problème des rapports pensée / langage / langues en éliminant toute distance entre modèle (description nécessairement linguistique d'une réalité cognitive qui n'a aucune raison *a priori* de l'être) et réalité (les processus de pensée). Cette façon de résoudre un problème en déclarant qu'il ne se pose pas ne permet ni de faire avancer la connaissance de la linguistique générale, puisque l'hypothèse du langage de la pensée n'a engendré, à ma connaissance, aucun programme de recherche concret portant sur l'apprentissage de langues particulières, ni de faire avancer la connaissance de la nature de la pensée. Bref, "l'hypothèse" en question est vouée, comme toute thèse métaphysique, et à la différence d'une hypothèse véritable, à mourir de sa belle mort : faute de combattants.

l'intelligence artificielle existe. Prenons des exemples en faisant varier les critères d'"intelligence". Critère de créativité : parmi disons le million de programmes de longueur variée déjà écrits, quel programme parvient à démontrer de lui-même un ou des théorèmes mathématiques intéressants *et* inconnus jusqu'alors ? *Aucun*<sup>35</sup>. Critère d'intentionnalité : quel programme a été écrit qui comprenne le sens des concepts qu'il manipule et crée de nouvelles catégories sémantiques ? *Aucun*. Critère cognitif : quel programme permet d'établir un théorème incontestable touchant une capacité cognitive quelconque de l'intelligence, animale, humaine ou artificielle ? *Aucun*<sup>36</sup>. Critère d'auto-évolution : quel langage de programmation a développé de lui-même de nouvelles fonctions syntaxiques de telle sorte qu'il permette de mieux rendre compte d'une modification dans l'interaction avec l'environnement ? *Aucun*. On peut objecter que même si, jusqu'à présent, les programmes que l'on a écrits ne remplissent pas les tâches que je viens d'énumérer rapidement, ils pourront le faire dans l'avenir (et d'ailleurs, j'ai été contraint à l'instant, après mes réponses négatives, d'utiliser des notes, ce qui montre que des progrès ont été faits et qu'il n'y a pas de raison de penser qu'il n'y en ait pas encore). Cependant, il faut reconnaître dans ce cas que l'adéquation des modèles à la *réalité* des processus de l'intelligence est de l'ordre du simple *souhait*. Dès lors, on est en droit de se poser la question de ce que l'on peut légitimement attendre de l'intelligence artificielle.

De l'ordre du souhait, donc. On peut en effet considérer qu'il s'agit d'une attitude frileuse que de voir dans les modèles d'intelligence artificielle de simples simulations et qu'il faut réussir à reproduire réellement les processus de la pensée. En général, ce souhait provient d'un présupposé : pour être considérée comme une science, toute connaissance doit pouvoir se réduire à

---

<sup>35</sup> On a réussi à reconstruire des théorèmes élémentaires. On pourrait imaginer un programme très puissant capable de produire des théorèmes plus compliqués. Mais serait-on alors capable d'une part de *reconnaître* qu'il s'agit de théorèmes et d'autre part, de sélectionner ceux qui seraient réellement intéressants ? Rien n'est moins sûr. On serait obligé de faire confiance au programme et on se retrouverait du même coup dans la situation peu mathématique de G. H. Hardy qui disait, quand il fut confronté aux théorèmes sans preuve de Ramanujan, retrouvés dans ses carnets : «Ces théorèmes doivent être vrais parce que personne n'aurait pu inventer des choses pareilles».

<sup>36</sup> Certains théorèmes ont été établis touchant des phénomènes cognitifs. Par exemple, les modèles de mémoire dans le cadre des réseaux de neurones ont établi le théorème suivant : il n'est pas contradictoire de construire un modèle dans lequel existe une mémoire distribuée. Mais cela ne veut pas dire que le réseau de neurones possède une mémoire *cognitive* et que le modèle soit pour autant devenu identique à la réalité. Cela veut dire seulement que le modèle dans lequel on use d'une mémoire distribuée apparaît comme une bonne approximation de ce qui peut se produire dans la nature (et peut-être même comme une meilleure approximation que celle accessible par un modèle à mémoire non distribuée).



un processus matériel qui soit isolable au sein de la réalité physique et dont le schéma causal soit mathématiquement descriptible. Dans cette optique, il n'y aurait pas d'objet scientifique qui soit d'un autre type que celui de la physique. Mais dans le cas de l'objet pensée, on a vu qu'il fallait prendre certaines précautions pour éviter de tomber dans le cercle vicieux qui consiste à vouloir établir la matérialité de la pensée puisque la description qu'on en fait n'est pas matérielle, à moins de présupposer que le modèle est adéquat à la réalité de l'objet, *ce qui est justement à prouver*.

À choisir entre un réalisme vantard et un constructionnisme frileux, il me paraît donc plus expédient de considérer jusqu'à nouvel ordre les modèles informatiques de l'intelligence artificielle selon le deuxième point de vue<sup>37</sup> et c'est dans cette optique que l'intelligence artificielle me semble être parvenue à transformer le sens de la notion d'expérience<sup>38</sup>.

### 3.1.2. La mathématisation

---

<sup>37</sup> Le terme de constructionnisme n'est d'ailleurs peut-être pas très heureux parce qu'il fait croire que l'attitude philosophique sous-jacente est un nominalisme qui voudrait se passer de toute prise de position ontologique. Or ce n'est pas le cas. Je pourrais prendre pour exemple du point de vue que je défends la façon dont, en physique, Bohr a énoncé le "principe de complémentarité" qui permet de sortir du dilemme touchant la nature corpusculaire ou ondulatoire des particules. En se débarrassant du réalisme classique qui tâchait en vain de régler une fois pour toutes la question de la nature des particules et en envisageant les deux points de vue concurrents comme des *modèles*, Bohr a adopté une attitude de prudence qui a libéré les esprits et a profité à la recherche. Cela ne veut d'ailleurs pas dire qu'il faille en rester là et que toute question touchant la nature réelle des particules soit à tout jamais bannie : la physique contemporaine a des moyens que n'avait pas celle de Bohr et qui permettent d'assurer aujourd'hui que les particules ne sont ni onde ni corpuscule, ce qui est un progrès capital d'un point de vue réaliste. La prudence épistémologique quant à la portée ontologique des modèles n'est donc pas nécessairement liée à une philosophie nominaliste.

<sup>38</sup> Il est sans doute possible de ne pas se laisser enfermer dans ce dilemme si l'on envisage les modèles de l'intelligence artificielle comme des expressions de l'organisme biologique que nous sommes. Une tâche épistémologique consisterait alors à tenter de montrer quels sont les schèmes sensori-moteurs à partir desquels l'intelligence artificielle interprétée comme un type d'expression de l'organisme a pu germer. C'est dans cette optique que J. Largeault faisait remarquer : «Si on définit l'être humain par la raison, la logique, à cause de ses affinités avec l'identité, acquiert une grande importance. Si on définit l'être humain par la vie, la physique ou la biologie gagnent le premier rang d'importance. Mais il n'est pas exclu, en dépit des apparences, que la logique soit une production du vivant. Le jour où nous saurons si, à quelque niveau, le vivant n'engendre pas quelque ordre analogue à celui de la logique ou de l'arithmétique récursive, nous serons mieux fixés sur sa signification réelle. Actuellement nous n'avons que la spéculation et nos partis pris». J. Largeault, (1990) : 13.

La mathématisation telle qu'on la rencontre en intelligence artificielle n'est ni une mathématisation au sens moderne telle qu'on l'entend depuis le XVII<sup>ème</sup> siècle et qui consiste en un engendrement mathématique de la réalité préalablement abstraite de ses "qualités secondes"<sup>39</sup>, ni une mathématisation au sens antique qui consiste en une application des mathématiques à la réalité perçue en vue d'en décrire les apparences. Elle est liée avant tout, comme je viens de le faire remarquer, à l'usage que l'intelligence artificielle fait de la notion de modèle cognitif. C'est pourquoi la mathématisation ne porte pas directement sur le monde mais seulement sur une version "apprivoisée" de nos rapports à celui-ci, limités à quelques traits saillants. Par exemple, dans le cas des réseaux connexionnistes de la mémoire tels qu'ils sont présentés par B. Victorri, on se limite à la mémorisation de trois formes-types et le réseau lui-même n'est composé que d'une vingtaine d'unités interconnectées, échantillon extrêmement faible quand on le compare au nombre des neurones réels du cerveau.

Cette version "apprivoisée" du monde est d'une nature *déjà abstraite* dans la mesure où le sujet connaissant est conçu comme faisant face à un monde déjà constitué sans lui, qu'il a les moyens d'appréhender à partir d'outils d'emblée représentationnels se présentant à lui soit sous forme logique, soit sous forme dynamique au sens mathématique du terme. Les questions proprement phénoménologiques de constitution des phénomènes par l'activité perceptive d'un sujet placé dans le monde sont donc laissées de côté.

La constitution du monde par les sujets est tout autre car nous ne rencontrons pas un monde déjà constitué d'objets stables sur lesquels nous pourrions directement accoler des noms qui les représenteraient. La perception est en effet un résultat qui s'opère par le biais d'une double maturation : celle du corps et celle de l'affect. À l'origine, chez les nourrissons, la perception est en effet un agent de survie qui est intégralement tourné vers la satisfaction des besoins vitaux. La fonction perceptive du corps débute dans la cavité orale (bouche, langue, lèvres, voies nasales et pharynx) et s'étend ensuite à la main, à la peau, au labyrinthe de l'oreille interne et à l'estomac<sup>40</sup>. L'aspect affectif de la perception provient de ce que c'est par l'intermédiaire du plaisir et du déplaisir que l'objet est perçu. Les objets n'ont donc de *signification* que par le biais de cette double détermination corporelle et affective : qu'ils assurent la survie de l'organisme en se manifestant à celui-ci sous l'aspect du plaisir. On trouve

---

<sup>39</sup> Les physiciens-mathématiciens du début du XVII<sup>ème</sup> siècle opposaient, comme on l'a vu au § 2.2, les "qualités premières" (figure, étendue et mouvement) aux qualités secondes (couleur, odeur, etc).

<sup>40</sup> Cf. R. A. Spitz, (1968) : 46-57.

des traces dans la langue naturelle du statut corporel et affectif lié à l'objet : Spitz a par exemple montré comment l'apprentissage du "non" et la gestuelle qui l'accompagne (dans la civilisation occidentale, le mouvement de la tête de droite à gauche) pouvait être décrite comme une transformation du schéma sensori-moteur de rotation de la tête utilisé par le nouveau-né à la recherche du sein de sa mère<sup>41</sup>.

Il y a donc des présupposés touchant la phénoménologie induite quand on a recours à la mathématisation en intelligence artificielle, présupposés qui dérivent de l'abstraction de la notion de modèle. Cette abstraction est particulièrement frappante dans le cas des modèles de l'intelligence artificielle classique où c'est le schème inférentiel logique qui doit rendre compte à lui tout seul de la structuration du monde. L'abstraction apparaît alors comme un défaut inhérent à ce type de mathématisation quand elle n'est pas conscientisée, c'est-à-dire quand n'a pas été précisé le rôle joué par la notion de modèle : elle tend en effet à véhiculer des présupposés naturalistes touchant la constitution du monde, comme l'avait déjà remarqué Husserl à propos de la mathématisation galiléenne<sup>42</sup>.

Cet objectivisme abstrait a également une contrepartie "subjective" : l'abstraction peut, dans certains cas qui restent à définir, occulter le *psychologisme* de la démarche du modélisateur. Il ne suffit pas en effet de constituer un modèle *mathématique* pour échapper aux illusions de l'introspection : précisément parce que les traits saillants du modèle sont déjà des traits abstraits, ils sont le résultat d'un processus mental visant l'abstraction, processus qui s'est effectué à l'extérieur de la description du modèle lui-même et souvent à l'insu du modélisateur. Or cette élaboration n'est pas toujours exempte de présupposés psychologisants dont la conséquence la plus visible consiste à considérer que la recherche de l'abstraction est la *garantie* de l'objectivité du modèle.

En effet, c'est souvent l'abstraction elle-même qui n'est justifiée qu'en arguant du fait que, pour que le modèle qui est, en somme, un "modèle réduit" du monde corresponde au monde réel, il suffirait qu'on augmente sa taille. Or il serait hasardeux de considérer que les modèles pourraient finir par s'identifier à la façon dont nous percevons le monde réel par simple accroissement quantitatif : rien ne dit en effet qu'un simple accroissement quantitatif des données prises en compte ne change pas radicalement la nature de la modélisation *parce que les changements d'échelle entraînent*

---

<sup>41</sup> Ibidem : 144-146.

<sup>42</sup> E. Husserl, (1935-1936) : 27-68.

*aussi des changements de description*<sup>43</sup>. Par exemple, dans des réseaux de neurones, quand on augmente le nombre des connexions entre neurones, ce n'est plus le langage discret du calcul sur des éléments individués qui est pertinent mais celui de la cinétique des gaz. L'abstraction doit donc être justifiée au coup par coup, sur chaque cas particulier, *pour que le modèle reste modèle*. Sinon, on tire hâtivement des conclusions sur la nature du monde ou de l'esprit, conclusions qui n'ont de valeur qu'autant qu'on accepte le présupposé de l'abstraction.

Prenons-en comme exemple la façon dont D. Sperber décrit la "révolution cognitive" en s'appuyant sur l'article de Turing fondant la théorie de la calculabilité, "On Computable Numbers with an Application to the Entscheidungsproblem"<sup>44</sup> :

«Cette nouvelle conception des rapports entre pensée et matière, comme on dit aujourd'hui, procède d'une découverte, mais il s'agit d'une découverte logique et non pas empirique. En 1936, alors que la logique mathématique était en plein essor, le mathématicien anglais Alan Turing a imaginé un dispositif matériel assez simple, connu depuis sous le nom de machine de Turing ; il a démontré qu'une telle machine serait capable en principe d'effectuer non seulement certains traitements d'information, des calculs par exemple, mais même tout traitement d'information que tout autre dispositif fini, un cerveau ou, aujourd'hui, un ordinateur par exemple, serait capable d'effectuer. En un mot : la découverte de Turing permettait de comprendre comment de la matière pouvait penser».

Cette description de l'article de Turing paraît quelque peu biaisée parce que D. Sperber opère comme un télescopage historique entre le point de vue de Turing qui fait intervenir le corps, l'écriture et le comportement et le point de vue, plus tardif, du fonctionnalisme qui fait intervenir un mentalisme, un solipsisme et la notion de symbole de la pensée. En effet, il n'est pas question dans l'article de 1936 d'une comparaison entre la "machine de Turing" et tout autre dispositif fini pour la bonne raison que la machine de Turing est une *machine infinie*, comme l'a définie d'emblée Turing<sup>45</sup>, qui ne

---

<sup>43</sup> B. Victorri va même plus loin quand il déclare dans (1995) : 339 : «[...] on peut même affirmer que ces petites maquettes sont sûrement plus intéressantes par elles-mêmes que ne le serait un système plus complet, construit de toutes pièces, qui aurait peu de chances de toutes façons de correspondre à la réalité [...]»

<sup>44</sup> D. Sperber, (1992) : 397-420. Le passage cité se trouve p. 400.

<sup>45</sup> Et à sa suite M. Davis que D. Sperber cite pourtant en note. Cf. M. Davis, (1958) : 3 : «Tandis que les machines à calculer physiques sont handicapées par le fait qu'elles n'ont accès qu'à une région finie pour conserver les données et les différentes expressions intermédiaires formées au cours du calcul (par exemple, les produits partiels lors d'une multiplication), nous imaginerons une machine à calculer qui imprime des symboles sur un ruban linéaire, en faisant l'hypothèse qu'il est *infini* dans les deux directions, divisé en une suite infinie de cases dans les deux sens». C'est parce que la machine de Turing est une machine infinie qu'elle permet

peut donc jamais être matérialisée. En revanche, la machine de Turing est bien une machine infinie dont la *table d'instructions* possède un nombre fini d'états (qui peuvent être égrenés indéfiniment) : il y a donc confusion dans la citation de D. Sperber entre la forme du dispositif logique et son contenu. C'est la raison pour laquelle il n'est nulle part question dans l'article de Turing d'une comparaison entre la notion de machine de Turing et ce dispositif fini particulier qu'est le cerveau et encore moins d'une comparaison entre pensée et matière car, comme le dit fort justement D. Sperber, l'article de Turing est un article de *logique mathématique* dans lequel la notion de matière n'aurait rien à faire. On assiste donc à une double assimilation dans cette citation : de la pensée à un dispositif fini — au sens combinatoire du terme — d'une part et de la pensée à de la matière d'autre part. La première assimilation ne se défend que si l'on peut montrer en quoi la pensée peut se comparer à un automate *fini*, ce qui n'est pas le but de l'article de Turing. La deuxième assimilation ne se défend que si l'on suppose que la pensée est matérielle, ce qui suppose le problème résolu, le modèle ne servant plus alors que de simple confirmation à ce présupposé. Or tout ce que peut faire le modèle, c'est *donner à penser* qu'il y a bien quelque chose de matériel dans l'effectuation de la pensée. Mais on n'a pas attendu le modèle informatique élaboré à partir du résultat de Turing pour s'en douter et il n'y a aucune raison de considérer que ce modèle apporte une preuve définitive ou même seulement supplémentaire au dossier de la matérialité de la pensée, sans doute ouvert par Platon il y a vingt-cinq siècles quand celui-ci énonça dans le *Théétète* trois hypothèses sur le fonctionnement de la pensée<sup>46</sup>.

Si l'on reprend en revanche l'article de Turing, on voit que celui-ci met en parallèle ce que fait l'être humain en train de calculer (ses "états d'esprit" et non sa pensée) et ce qu'il est possible d'en traduire par l'écriture<sup>47</sup> (c'est cette traduction écrite qui s'appelle "machine", pour Turing). L'attention de Turing se porte donc sur la traduction symbolique d'un état d'esprit, la seule véritable réalité étant finalement ce qu'il est possible de décrire par le *langage* d'un point de vue *fini*, ce qui n'est pas un point de vue mentaliste : il s'agit seulement de la description d'un couplage entre une mécanique calculante et un milieu d'opérations lui-même déjà discrétisé. Plus

---

de définir la classe des nombres réels calculables : dans l'article de Turing, une machine de Turing est considérée comme correctement définie *si et seulement si* elle continue à calculer indéfiniment l'approximation décimale du nombre réel qu'elle permet d'approcher. Sinon, il s'agit d'une machine appelée par Turing "circulaire". cf. A. M. Turing, (1936) : 230-265, § 8.

<sup>46</sup> En faisant cette dernière remarque, je ne veux pas donner l'impression de nier les acquis des quarante dernières années mais seulement de préciser ce que l'on peut attendre d'un modèle cognitif : non pas une corroboration de la matérialité de la pensée — point de vue théologique inversé — mais une simulation d'un certain nombre de processus qu'on peut manifester matériellement et qui sont dans un rapport d'analogie avec des processus que l'on attribue à la pensée.

<sup>47</sup> Cf. A. M. Turing, (1936), § 1.

profondément, Turing met l'accent sur le rôle actif du langage dans la constitution des états mentaux, c'est-à-dire sur le fait que *c'est lui qui façonne nos états mentaux* : en effet, ce sont finalement les symboles écrits qui nous mettent dans un certain type d'état mental (le type d'état mental de l'être humain en train de calculer) et non l'inverse. Une conséquence majeure découle de cette analyse : la réalité des états mentaux "en eux-mêmes" est mise entre parenthèses puisqu'on ne les appréhende qu'une fois traduits, c'est-à-dire déjà symbolisés par le langage. Il n'y a donc aucune prise de position philosophique de la part de Turing quant à la *nature* des états d'esprit, matériels, immatériels, finis ou infinis, purement intérieurs au sujet ou non et *c'est précisément dans cette mesure* qu'il est possible de construire un *modèle* de l'esprit à partir de son travail de logique mathématique<sup>48</sup>. Sinon, on finit par confondre l'abstraction logique que représente la machine et la réalité de l'esprit et par conclure que l'esprit *est* l'abstraction qui nous sert à le décrire. On en vient à conclure que l'abstraction est la matière du cerveau, ce qui est la marque même du *psychologisme* qui confond le point de vue réel du fait et le point de vue normatif du droit<sup>49</sup>. C'est bien parce que Turing, en ayant recours à la notion de machine infinie, ne se prête pas à cette confusion qu'il est possible de constituer un modèle *mathématique* de la notion *psychologique* d'état de l'esprit.

### 3.1.3. L'instrumentation

On a vu tout à l'heure que l'une des transformations héritée de la science classique du XVII<sup>ème</sup> siècle était d'avoir élaboré une théorie du corps-machine conçue comme un choc en retour de la constitution mathématique du corps en instrument. Il y a un aspect paradoxal dans cette constitution "en retour" du corps en instrument parce qu'elle est non-intuitive et qu'elle se fonde sur la distinction entre le corps vécu et le corps abstraitement pensé. C'est le même paradoxe qui a beaucoup fait pour discréditer au XX<sup>ème</sup> siècle le projet de l'intelligence artificielle et nourrir pendant longtemps la prévention du public philosophique.

---

<sup>48</sup> Cela ne veut pas dire que l'analyse de l'acte de calcul telle qu'elle est menée par Turing soit exempte de prise de position philosophique : au contraire, l'analyse de Turing est *foncièrement nominaliste*, dans la mesure où elle privilégie le réalisme des individus, d'une part en envisageant le langage lui-même du point de vue de la matérialité individuelle de ses marques et d'autre part en considérant comme seuls réels des *états* discrétisés de l'esprit.

<sup>49</sup> J'utilise ici la différence juridique entre l'état de fait et ce qui le rend légitime. Par exemple, on oppose en droit la *possession* qui relève du fait à la *propriété* qui rend cette possession légitime au regard de la loi.

### 3.1.3.1. Remarque sur l'interrogation philosophique touchant la légitimité de la perspective de l'intelligence artificielle

La question des philosophes, plus ou moins bien explicitée, est généralement la suivante : comment est-il possible de constituer *en retour* une analyse mathématique de l'activité de pensée, c'est-à-dire comment pourrait-on envisager de l'extérieur l'activité de pensée alors même que c'est d'elle que procède toute la constitution mathématique du monde ? C'est cette interrogation qui, pour Husserl, rendait obligatoire, sous peine de sombrer dans le psychologisme, une analyse de la conscience en termes phénoménologiques. L'intelligence artificielle a définitivement montré qu'il était possible *d'éviter l'écueil du psychologisme par d'autres voies que la voie purement phénoménologique* — même si, pendant sa période classique, elle s'est souvent complue dans le psychologisme -. Mais *elle n'en a pas résolu pour autant le problème central de la phénoménologie*, celui de la constitution du monde : pour comprendre cette problématique à deux temps, il faut comprendre que l'intelligence artificielle n'est *pas encore* une physique de l'esprit car elle n'est qu'une *simulation* des processus de pensée.

Commençons par examiner une des réponses provenant de l'intelligence artificielle au problème du psychologisme. Celle-ci est la suivante : il est possible de constituer une analyse mathématique de la pensée car on a les moyens mathématiques d'en simuler les processus, c'est-à-dire de faire des *modèles* de ces processus. Ces modèles sont des modèles de simulation qui n'ont pas la prétention d'être prédictifs : ils ne visent donc ni à reproduire ni à recréer les processus de pensée, mais seulement à en étudier les modalités idéales d'effectuation. Il n'y a donc pas de risque particulier de sombrer dans le psychologisme en élaborant ces modèles, tout au moins pas plus que dans la perspective phénoménologique elle-même dont l'outil principal, celui de la réduction phénoménologique, pourrait bien n'être finalement qu'une forme sophistiquée de psychologisme retranscrite dans un vocabulaire différent, comme Piaget l'avait noté en son temps<sup>50</sup>. Reste le problème de la constitution mathématique du monde.

Si l'intelligence artificielle était une véritable physique de l'esprit, elle serait en mesure de décrire mathématiquement le processus de constitution du monde par la pensée. Il lui faudrait occuper une position d'extériorité par rapport à la pensée, ce qui lui permettrait d'élaborer

---

<sup>50</sup> Cf. J. Piaget, (1965) : 148 : «[...] la phénoménologie n'est qu'un psychologisme comme un autre, mais développé en un autre langage».

des modèles globalement ou localement prédictifs. Mais *elle est incapable de le faire*, non pas parce qu'elle serait une science moins avancée que la physique mais parce que la pensée n'est pas un objet de même nature que les objets matériels dont s'occupe la physique. Bref, c'est la spécificité de son "objet", *intentionnel* avant d'être physique, qui exige de la part de l'intelligence artificielle une méthodologie modélisatrice tout aussi spécifique. Tout tourne donc autour de l'interprétation du terme de modèle et de la pertinence plus ou moins réaliste que l'on accorde à ce terme, comme je l'ai déjà noté.

### 3.1.3.2. L'instrumentation du cognitif

Remarquons tout d'abord que l'ordinateur et les langages de programmation n'ont été indispensables ni à la constitution du projet d'une connaissance scientifique de l'esprit dans le passé<sup>51</sup> ni à sa bonne marche jusqu'à aujourd'hui<sup>52</sup>. L'ordinateur ne semble être, de ce point de vue, qu'un *outil* — il est vrai peut-être primordial — dans le cadre des recherches en intelligence artificielle. Il y aurait donc un malentendu à vouloir constituer le projet de l'intelligence artificielle autour de la seule idée d'une application de mécanismes aux processus intelligents. C'est la remarque critique que David Marr faisait à propos d'un certain nombre de programmes d'intelligence artificielle<sup>53</sup> :

«Si l'on croit que le but des études de traitement de l'information est de formuler et de comprendre des problèmes particuliers de traitement de l'information, alors c'est la structure de ces problèmes qui est centrale, non pas les mécanismes au moyen desquels ils sont incarnés [...]. On peut faire une critique [...] des travaux de Norman et Rumelhart dans lesquels les auteurs étudient la façon dont l'information semble être organisée dans la mémoire à long terme. Encore une fois, le danger vient de ce que les questions ne sont pas posées en rapport à un problème de traitement de l'information qui soit bien défini. Au lieu de cela, elles sont posées, ainsi que les réponses proposées, en termes de mécanisme — en l'occurrence, le mécanisme est appelé un "réseau structural actif" et il est si simple et si général qu'il est dénué de toute substance théorique. Les auteurs sont peut-être capables de dire que telle ou telle "association" semble exister mais ils ne peuvent pas dire en quoi consiste cette association, ni qu'elle doit être ainsi parce que, pour résoudre un problème X (que nous pouvons résoudre),

<sup>51</sup> Comme le fait remarquer J.-P. Dupuy, le projet de constitution d'une "physique de l'esprit" a été conçu avant l'utilisation effective des ordinateurs et n'en a donc pas été le moteur. Cf. J.-P. Dupuy, (1994).

<sup>52</sup> D. Andler parle à propos de l'usage de l'ordinateur d'une «abstinence méthodologique» prônée par certains chercheurs. Cf. D. Andler, (1990) : 69. Cf. aussi, B. Chandrasekaran, (1990).

<sup>53</sup> D. Marr, (1977) : 37-48.



on a *besoin* d'une mémoire organisée de telle ou telle manière et que si on l'a, certaines associations apparentes se manifestent comme effets secondaires.»

Bref, ce n'est pas à partir d'une *application* de la notion de mécanisme à certains processus apparaissant dans des tâches réputées intelligentes que l'intelligence artificielle procède : c'est en isolant des fonctions sous forme de problèmes dont elle essaye d'imaginer théoriquement la structure calculable — comme par exemple le problème de la détermination du système de coordonnées locales permettant la détection des contours d'une image — que l'intelligence artificielle en vient à décrire, dans un deuxième temps, ce que pourrait être d'une part l'implémentation de cette structure dans un langage de programmation et d'autre part la réalisation matérielle d'une telle structure dans un *mécanisme* de traitement de l'information. Il y a donc une tâche à accomplir avant de s'intéresser aux mécanismes de l'intelligence proprement dits : cette tâche est celle d'un *détour par le théorique* qui ressemble fort à la façon dont, en physique, l'instrumentation depuis Galilée consistait à déterminer le rapport mathématique entre une fonction et sa matérialisation.

Or si l'on s'en tient à une simple description en termes mécaniques, on finit, comme dit D. Marr, «par aboutir à des mécanismes improbables dont la seule caractéristique est qu'ils ne peuvent pas faire ce que nous ne pouvons pas faire». Ces mécanismes, dont le statut épistémologique est peu déterminé, ont donc un pouvoir explicatif à peu près nul touchant la nature du fonctionnement cognitif. Dans la mesure où ils ne donnent pas les raisons pour lesquelles telle ou telle tâche est accomplie de telle ou telle manière, ils restent purement contingents et on peut les considérer comme faisant partie de cette version contemporaine et formalisée du *psychologisme*, version dont on a déjà vu la manifestation quand on a abordé le thème de la mathématisation.

Le seul moyen d'éviter le travers du psychologisme qui guette celui qui tente de mettre au jour le fonctionnement d'un processus attribué au mental est de tenter de tracer la limite entre ce qui peut être accompli et ce qui ne peut pas l'être par rapport aux conditions imposées théoriquement par le problème. On peut faire, de ce point de vue, une analogie avec le projet élaboré par Turing à partir de son théorème de limitation<sup>54</sup> : c'est à partir du tracé de la limite entre calculable et

---

<sup>54</sup> Par théorème de limitation, je veux faire allusion à la solution négative au problème de l'arrêt.

incalculable pour des classes données de problèmes qu'il peut circonscrire négativement ce qui relève du calculable ; de même ici, c'est en réussissant à formuler une question d'ordre théorique touchant la limite entre ce qui relève du calculable et ce qui n'en relève pas que l'on peut se demander, dans le cas où une classe de problèmes calculables a été isolée, si une solution calculable pourra valoir comme réponse à ce problème, avant même de savoir si des algorithmes sont ou non à notre disposition. Il est possible qu'il y ait une réponse en termes calculables tandis qu'il y a des questions dont on ne sait pas si elles ont ou auront jamais une réponse en termes calculables<sup>55</sup>.

On doit donc considérer la démarche de la modélisation en intelligence artificielle comme incluant deux étapes bien distinctes : une étape abstraite au cours de laquelle on doit poser un problème par rapport à une fonction déterminée ; une étape plus technique dans laquelle on doit se demander *comment* peut s'incarner dans un mécanisme de type algorithmique la fonction décrite en termes calculables.

Ces points méthodologiques éclaircis, revenons maintenant aux cinq traits, globaux et locaux, dont nous nous étions servi en commençant pour caractériser la notion d'expérience.

### **3.2. Les traits globaux de la notion d'expérience et l'intelligence artificielle**

Ils sont deux : la *cohérence* du domaine d'événements et son *ouverture* indéfinie.

#### *3.2.1. Cohérence du domaine d'événements propres aux modèles informatiques : l'écriture*

Comment définir la cohérence du domaine d'événements dans le cadre de l'intelligence artificielle ?

La différence majeure qui existe entre les événements du domaine de référence et les événements du modèle en tant qu'implémentation informatique rendant compte du domaine de référence est de nature temporelle : le temps dans le domaine de référence est *irréversible* alors que le temps du modèle est, lui, *réitérable*. C'est ce qui en fait

---

<sup>55</sup> Dans le vocabulaire de D. Marr, les problèmes que l'on peut considérer comme calculables sont des problèmes de "Type-I"; c'est le cas, pour lui, de la théorie chomskienne de la compétence pour la syntaxe de l'anglais. Les problèmes qui ne sont pas calculables, problèmes de "Type-II", sont trop complexes pour recevoir une solution en termes calculables; le problème gestaltiste "figure-fond" lui semble être l'exemple d'un problème de type II.

d'ailleurs une réalité virtuelle : le temps du réel y est indéfiniment différé. À quoi tient l'aspect réitérable du temps qui s'inscrit dans le modèle ? Au fait que la cohérence du domaine d'événements constituant le champ de l'intelligence artificielle exige que tout événement apparaisse sous l'aspect d'un *jeu d'écriture*<sup>56</sup>.

Si l'on reprend l'exemple de la description de l'acte de calcul telle qu'elle est proposée par Turing dans son article *princeps* de 1936, on voit que chaque case du ruban d'une machine de Turing n'a le sens d'une entrée qu'en tant qu'il est rempli par un trait — ou son absence — interprété comme *signe* et que la table d'instructions prescrivant un comportement pour chaque configuration dans laquelle la machine peut se trouver (c'est-à-dire le couple formé de l'état interne de la machine à un moment *t* et de la case observée) joue le rôle de *grammaire*. On pourrait objecter que l'intelligence artificielle a précisément montré par son relatif "échec" que le schéma d'inférence (axiomes + règles d'inférence) tel qu'il a été explicité par Turing n'était pas suffisant pour rendre compte d'un comportement intelligent, parce qu'un comportement fait appel à des connaissances implicites qui sont peut être d'une autre nature que celles que l'intelligence artificielle est en mesure de reconnaître (les rôles du corps et de "l'être-au-monde" en feraient partie). Mais l'intelligence artificielle ne fait pas intervenir pour autant autre chose qu'un jeu d'écriture constitué autour du schéma d'inférence : ce qui change, c'est le rôle des connaissances au sein du schéma.

Aussi une réalité ne devient-elle objet d'expérience en intelligence artificielle que si elle possède une structure de signe réinscriptible dans le médium informatique. Cette condition de cohérence n'est pas à proprement parler nouvelle dans l'histoire de la pensée. C'était déjà le point de vue de Leibniz, si l'on en croit Couturat qui faisait remarquer que, pour ce dernier<sup>57</sup>,

«la nature est le produit d'une logique divine, de ce calcul immense qui est la création ; elle est pour nous une admirable machine à calculer, car elle nous fournit, tout faits, les résultats de calculs qui dépassent la portée de notre entendement».

L'intelligence artificielle a cependant eu le mérite de rendre ce point de vue effectif en parvenant à le séparer du point de vue de la science

---

<sup>56</sup> C'est ainsi que Turing définissait déjà la notion de "mécanique" dans (1950) : 456 : «mécanisme et écriture sont de mon point de vue presque synonymes».

<sup>57</sup> L. Couturat, (1901) : 256. Leibniz est sûrement l'auteur de la tradition la plus proche de la perspective adoptée par l'intelligence artificielle classique.

physique dont les objets ne sont pas soumis à la même condition de cohérence puisque les entités étudiées n'ont pas nécessairement l'aspect de signes. C'est cette condition de cohérence qui constitue la possibilité même de l'aspect fonctionnel de la méthode utilisée dans les modèles jusqu'ici construits en intelligence artificielle.

On peut donc dire que la cohérence du domaine d'événements propre à l'intelligence artificielle exige l'élaboration d'un niveau spécifique d'intelligibilité, celui d'un monde déjà sémiotiquement retravaillé. Le paradigme de cette cohérence, c'est le *jeu*. C'est d'ailleurs la notion de jeu qui a servi de modèle d'investigation en intelligence artificielle dès les années 50<sup>58</sup>.

### 3.2.2. Ouverture indéfinie du domaine des événements informatiques : le discret comme milieu de donation

Le second trait caractéristique de la notion d'expérience consiste à reconnaître que la cohérence du domaine d'événements n'empêchait pas l'*ouverture* indéfinie du domaine : les événements apparaissant au sein du domaine sont repérables quoiqu'en nombre indéfini.

Dans le cas de l'intelligence artificielle, l'ouverture du domaine en tant qu'il est modélisé informatiquement exige que l'on ait affaire à des événements discrétisés ou discrétisables. C'est cette caractéristique qui leur permet d'acquérir un statut de signe dont il est possible de rendre compte par le biais d'un calcul. Le type d'infini définissant l'ensemble des entiers naturels est donc requis pour rendre compte de l'ouverture du domaine d'expérience propre à l'intelligence artificielle. C'est la différence, du point de vue du calcul, entre ce type d'infini et un type d'infini supérieur qui permet de tracer la limite entre ce qui relève du domaine d'expérience de l'intelligence artificielle et ce qui n'en relève pas : la classe des nombres réels calculables est le résultat du tracé de cette limite. Du point de vue du calcul, l'intelligence artificielle entretient donc des rapports avec un domaine numérique qui est plus étendu que le sien (le domaine des nombres réels), domaine qu'elle doit en même temps présupposer dans son "ontologie" numérique et nier dans sa méthode. De ce point de vue, l'intelligence artificielle se trouve comme hantée par le continu à la fois d'un point de vue mathématique puisque ce qui constitue numériquement son domaine d'expérience se définit par rapport à lui et d'un point de vue physique puisque ce qui fait événement doit pouvoir revêtir une forme discrète, quelle que soit la

---

<sup>58</sup> Cf. D. Michie, (1974). D. Michie rapporte que c'est en 1959 que, pour la première fois, un ordinateur bat l'inventeur de son programme dans plus de 50% des cas (il s'agit du programme de jeu de dames conçu par A. Samuel).

réalité ultime des composants de la matière, si composants il y a. Le discret agit donc comme une contrainte qui rend l'expérience en intelligence artificielle possible.

On pourrait critiquer ce point de vue en faisant remarquer que la différence entre le continu et le discret n'est pas pertinente pour l'intelligence artificielle parce que seul le discrétisable serait pour elle porteuse d'objectivité. Je ne partage pas cet avis. La prévention à l'égard du continu provient de présupposés philosophiques matérialistes et atomistes qui s'énoncent de la manière suivante : la matière ultime étant granulaire et discrète, tout ce qui relève du continu n'est au mieux qu'une façon de parler et au pire qu'une illusion tenace chez quelques attardés. Pourtant cette interprétation est un présupposé philosophique et non une théorie scientifique. Jusqu'à preuve du contraire, discret et continu doivent s'opposer pour que le discret ait un sens.

### 3.3. Les traits locaux de l'expérience et l'intelligence artificielle

Ces traits, au nombre de trois, concernent la nature des événements qui, pour faire l'objet d'une expérience, doivent être *observables* et *réitérables* d'une part et peuvent être *inattendus* d'autre part.

#### 3.3.1. Aspect observable : principe de détermination

L'aspect observable des événements se produisant dans le domaine de l'intelligence artificielle exige que la détermination des événements se fasse selon un schème logique pour le moment principalement du type du calcul des prédicats (pris ici en un sens très général, non exclusif de la variété proliférante des logiques actuelles) ou que cette détermination soit le résultat stabilisé d'un processus comme dans le cas du schème dynamique. Cette détermination des événements apparente l'intelligence artificielle aux sciences dures et lui assure une méthodologie très sûre, rendant possible une accumulation progressive de résultats. C'est ce que faisait remarquer D. Marr dans le cas de l'intelligence artificielle classique<sup>59</sup> :

«A proprement parler, un *résultat* d'intelligence artificielle consiste dans le fait d'isoler un problème de traitement d'information particulier, de formuler une théorie informatique pour lui, de construire un algorithme qui l'incarne et d'obtenir une démonstration pratique que l'algorithme est efficace. Le point important à souligner ici est qu'une fois qu'une théorie informatique a été établie pour un problème particulier, elle n'a jamais à être refaite et c'est à cet égard qu'un résultat d'intelligence artificielle se comporte de la même manière qu'un résultat en mathématique ou dans n'importe quelle autre science naturelle dure.»

---

<sup>59</sup> D. Marr, (1977) : 37-48.

Aussi un résultat informatique, une fois déterminé, est-il acquis et n'a plus besoin d'être reconfirmé (sauf s'il est retrouvé par d'autres biais) : l'aspect déterminé des résultats permet l'accumulation.

### *3.3.2. Aspect réitérable : principe de reproduction*

Pour faire partie d'une expérience, un événement doit pouvoir se reproduire : il n'y a pas d'événement absolument unique parce que cela remettrait en question les traits globaux de la notion d'expérience, la cohérence et l'ouverture indéfinie du domaine. Cette possibilité d'une reproduction des événements rend possible l'un des buts que s'est assignée l'intelligence artificielle, l'apprentissage, lié dès l'origine au projet d'intelligence artificielle : pour reprendre le cas de l'article de Turing de 1950, toute la dernière partie de l'article porte sur la question de l'apprentissage<sup>60</sup>. Turing fait intervenir plusieurs éléments pour élucider la question de la nature de l'apprentissage par une machine. Tout d'abord, un facteur externe, à savoir le jugement du programmeur touchant l'efficacité de son programme : le programmeur peut ajuster des paramètres pour rendre le programme plus adéquat à la tâche qui lui est assignée. Ensuite un facteur interne, à savoir la transformation par la machine elle-même de ses propres règles, soit par le biais déterministe d'une transformation des critères de pertinence touchant l'application des règles, soit par le biais d'une marge aléatoire dans leur application. On voit donc déjà se profiler les différentes façons dont sera abordé par la suite le problème très complexe de la représentation des connaissances (problème de l'inférence sur les connaissances acquises et problème du raisonnement plausible laissant une part à l'aléatoire).

### *3.3.3. Aspect inattendu : principe d'inspection*

L'inattendu en intelligence artificielle est, comme dans toute science, à la fois susceptible de se produire et néanmoins toujours transitoire. Sa détermination passe par l'articulation de différentes échelles temporelles. Il faut en effet distinguer deux échelles dans l'apparition des événements, l'échelle de l'effectuation et celle du contrôle. Du point de vue de l'effectuation, il est impossible, au moment où un programme est exécuté par un ordinateur ou qu'un réseau de neurones est en activité, de se situer du point de vue de ce qui se produit dans les langages de plus bas niveaux ou de l'activation des neurones : les événements sont en trop grande quantité et apparaissent de façon trop rapide pour qu'ils puissent faire sens pour un observateur. Néanmoins, du point de vue du contrôle, tout événement doit pouvoir

---

<sup>60</sup> Cf. A. M. Turing, (1950), § 7.

être explicité dans un temps ultérieur, que ce soit du point de vue du langage de programmation ou du point de vue de l'occupation de l'espace du réseau.

Ainsi l'inattendu se soumet-il à ce que l'on pourrait appeler le principe *d'inspection différée*, principalement illimitée et exhaustive : même si en fait les événements sont obscurs, en droit, ils peuvent toujours être éclaircis et *isolés* les uns des autres. Un événement qui ne serait pas susceptible d'être isolé — ne serait-ce que dans une région de l'espace du réseau — ne serait pas un événement d'intelligence artificielle parce qu'il resterait inattendu et partant, indéterminable.

### **3.4. L'intelligence artificielle et l'évolution de la notion d'expérience**

Il est maintenant possible d'intégrer la perspective ouverte par l'intelligence artificielle à l'évolution historique de la notion d'expérience. Sans marquer une véritable révolution touchant l'interprétation de cette notion, il faut noter cependant que l'intelligence artificielle a mis l'accent de façon inédite sur un certain nombre de ses traits caractéristiques. Nous nous limiterons à deux d'entre eux qui sont intimement liés : la façon dont l'intelligence artificielle a profondément renouvelé la notion d'expérience de *pensée* d'une part et le nouveau traitement que l'intelligence artificielle réserve à la notion d'*inconscient* d'autre part.

#### *3.4.1. L'expérience de pensée*

Comme on l'a déjà souligné, la notion d'expérience de pensée confère à l'expérience une composante imaginaire qui ne l'empêche pas d'être interprétable en termes de vrai et de faux. Seulement les expériences de pensée telles qu'elles étaient imaginées par Galilée ont fini par être réalisées et restent, quoi qu'il en soit, réalisables dans l'extériorité de la nature. C'est en effet le recours à l'extériorité de la nature qui permet d'élaborer des contrôles grâce à des montages spécifiques. Or l'aspect nouveau de l'expérience de pensée dans le contexte de l'intelligence artificielle, c'est que certaines expériences de pensée ne peuvent pas être contrôlées dans l'extériorité de la nature ou déterminées par une mesure parce qu'elles ne visent aucune extériorité physique, alors même qu'elles font de celle-ci le substrat "constituant" de leur cours. Si donc un contrôle demeure possible, c'est que la notion d'extériorité revêt un autre sens que le sens traditionnel : il devient possible de concevoir une extériorité qui ne soit pas rapportée à la nature physique mais bien à la pensée elle-même.

Prenons comme indice de ce nouvel état de fait la façon dont A. Burks décrit le double usage de l'ordinateur, qu'il appelle "usage heuristique" et "usage de simulation", c'est-à-dire, selon nos conventions de vocabulaire, en tant qu'instrument et en tant qu'outil de calcul<sup>61</sup> :

«L'usage heuristique des ordinateurs est semblable à et peut être combiné avec la méthode expérimentale hypothético-déductive traditionnelle en science. Dans cette méthode, on fait une hypothèse sur la base de l'information accessible, on teste expérimentalement les conséquences et on forme une nouvelle hypothèse sur la base de ce que l'on a trouvé ; cet enchaînement est itéré indéfiniment. En utilisant un ordinateur de façon heuristique, on procède de la même manière, sauf que le calcul remplace l'expérience. On fait une hypothèse touchant les équations que l'on étudie, on essaye d'isoler certains cas spéciaux importants, on utilise l'ordinateur pour résoudre ces cas, on contrôle l'hypothèse au vu des résultats, on forme une nouvelle hypothèse et on réitère le cycle. Les calculs peuvent aussi être comparés avec des données expérimentales. Quand cela se produit, l'usage heuristique devient simulation.»

Ainsi y aurait-il des paliers d'intelligibilité successifs dans la notion d'expérience en intelligence artificielle, paliers qui iraient d'un rapport de *simulation* de l'extériorité de la nature à un rapport d'*investigation* de l'extériorité abstraite de la pensée. Ce serait finalement, dans le cas de ce dernier palier, l'aspect mécanique de la pensée qui jouerait le rôle d'une extériorité par rapport à la pensée intuitive et non-formelle<sup>62</sup>. L'intelligence artificielle aurait pour tâche d'étudier les propriétés de cette extériorité de la pensée par rapport à elle-même en se posant la question suivante : quelles sont les conséquences observables quand on laisse aller le dispositif mécanique de la pensée à sa spontanéité ?

De ce point de vue, l'intelligence artificielle a une place tout à fait spécifique dans l'investigation touchant la pensée, dans la mesure où elle accorde à la pensée abstraite, quand celle-ci est caractérisée comme mécanique, une générativité autonome vis-à-vis de la diversité matérielle du monde<sup>63</sup>, qui s'apparente ainsi à une *spontanéité*. Alors que la spontanéité est généralement associée au non-abstrait, à l'intuitif, et caractérise la proximité vivante de la pensée à elle-même, il y aurait aussi une autre spontanéité, ou encore une sorte de "liberté mécanisable" de la pensée, qui l'affecterait en même temps comme une

---

<sup>61</sup> J. Von Neumann, (1966) : 14.

<sup>62</sup> C'est sans doute vrai en mathématiques depuis la constitution par Hilbert d'une méta-mathématique dont le rôle est le contrôle finitaire des énoncés mathématiques. Cf. par exemple, D. Hilbert, (1923) : 131-144.

<sup>63</sup> Hormis bien sûr la matière déjà formatée de la machine.



extériorité. Ces deux spontanéités seraient susceptibles de se rencontrer et c'est précisément là que se tiendrait l'intelligence artificielle, faisant du point de passage entre parcours informels et procédures formelles l'objet même de sa recherche<sup>64</sup>.

Si l'on se sert de la citation de A. Burks comme d'un exemple, on remarque que l'usage de l'ordinateur en tant qu'outil de *calcul* permettant la simulation est présenté de façon annexe par rapport à son usage en tant qu'instrument de *calculabilité*. En effet, la notion d'automate abstrait change le rapport au nombre puisqu'elle permet de dresser une carte topographique de ce qui est accessible, inaccessible et relativement accessible dans le domaine du nombre en introduisant une mesure objective du degré d'accessibilité du nombre (sa complexité). C'est ce que note J. Hartong quand il fait remarquer<sup>65</sup> :

«La mathématique héritée des anciens grecs classait les nombres selon leur nature : entier, rationnel, irrationnel, comme l'astronomie disposait les astres sur une sphère. La mathématique à l'ère de l'ordinateur (conformément à la vision du génial précurseur Borel) perçoit les nombres dans une perspective où la distance est le degré de complexité, tout comme l'astronomie à l'ère du télescope a révélé que le firmament avait une profondeur et que la majorité des astres sont trop obscurs et éloignés pour être visibles.»

Aussi l'intelligence artificielle a-t-elle montré qu'il était possible de dissocier, dans la notion d'expérience, l'idée de *contrôle par une extériorité* et l'idée d'*extériorité de la nature*. La nature n'est donc pas la seule à posséder à la fois le dynamisme et la cohérence propres à "juger" nos jugements. Une autre possibilité s'offre quand l'être humain se place dans une situation où il remplace ses actions par des opérations et substitue à la multitude de ses intentions en acte une intention préalable de type systémique qui, elle, est mécanisable. Le rapport de la pensée à elle-même conçue sur le mode de l'extériorité se manifeste alors dans le décalage entre l'intention vivante et l'intention morte stockée dans le programme : le temps, différencié et réitérable, agit comme producteur d'extériorité pour la pensée.

Comment l'intelligence artificielle est-elle parvenue à dissocier contrôle par une extériorité et extériorité de la nature ? Pour autant qu'il soit possible d'en juger, cette dissociation n'est pas le fait de l'intelligence artificielle elle-même mais procède d'un changement en profondeur de l'interprétation du rapport entre les mathématiques et le

<sup>64</sup> C'est cette question du rapport du formel et de l'informel qui divise les philosophes de l'esprit quant aux buts réels poursuivis par l'intelligence artificielle. Cf. Pour un panorama général des thèses défendues, R. Penrose, (1994), chap. 1.

<sup>65</sup> J. Hartong, "Le continu et l'ordinateur", (1987) : 13-27.

monde physique, changement qui date du début du siècle au moins. On peut en effet conjecturer que l'avènement préalable des géométries non-euclidiennes a eu alors entre autres conséquences de montrer que la référence à l'extériorité de la nature comme fondement de l'intuition de l'espace (euclidien) était partielle et que le concept d'espace contenait plus que ce que la géométrie (empirique) d'Euclide y avait trouvé. Dès lors l'extériorité de la nature ne sert plus de caution physique au concept d'espace qui peut être conçu indépendamment d'elle. Il devient alors possible de concevoir l'extériorité — spatiale ou non — sur le mode d'un rapport *interne à la pensée*. C'est sûrement la théorie des ensembles qui a le plus contribué à ce changement d'attitude en définissant des opérations d'une *arithmétique* appropriée (fondée sur l'opération de mise à la puissance d'un ensemble) qui rendent actuel un *excès* de la pensée sur elle-même<sup>66</sup>, le problème devenant alors celui de la limitation de la taille des ensembles, version réaliste des questions touchant à l'accessibilité des nombres<sup>67</sup>.

Ce court aperçu historique laisse cependant entière la question de savoir comment concevoir, sans paradoxe, l'idée d'une extériorité *interne* à la pensée. Il faut, pour en rendre compte, avoir recours à la notion d'inconscient, dans une forme spécifique.

#### 3.4.2. *Légitimité d'une forme procédurale de la notion d'inconscient*

La conséquence majeure de la transformation du statut de l'extériorité est qu'*il peut y avoir quelque chose dans la pensée qui lui échappe partiellement*. Ce sont les modalités de ce qui échappe à la pensée, dans le cas spécifique de l'intelligence artificielle, qu'il faut maintenant essayer d'explicitier. Il faut, pour ce faire, bien prendre la mesure de ce que l'on entend par machine.

Il faut en effet prendre conscience de l'amphibologie qui existe dans l'usage du terme de "machine", tantôt pris abstraitement quand on parle de "machine de Turing" ou plus généralement d'automate abstrait, tantôt pris physiquement quand on pense aux machines qui nous entourent et qui sont conçues en vue de réaliser des tâches pratiques. Cette amphibologie ne doit pas faire oublier quatre points capitaux

---

<sup>66</sup> Cette question est traitée en profondeur dans J.-M. Salanskis (1991) : 100 : «Le développement de la théorie des ensembles vers la théorie des cardinaux souligne cet élément de "transcendance" : la preuve de l'excès de  $P(X)$  sur  $X$ , c'est-à-dire de ce que, dans l'ordre de l'infini, il existe un opérateur de dépassement toujours applicable, conduisant l'excès d'un infini donné au-delà de lui-même, manifeste [...]».

<sup>67</sup> Cf. M. Hallet, (1984) : 86 sq.

concernant la nature des machines informatiques. Premièrement, *elles ne sont pas des entités du monde de la nature* mais des dispositifs d'écriture et ne possèdent ni l'indépendance des objets ni l'autonomie des organismes. Deuxièmement, elles sont une certaine *forme d'expression* de la spontanéité de la pensée propre à l'être humain, exactement la formalisation partielle de son intuition mathématique, puisqu'elles en sont un formalisme abstrait : dans cette mesure, elles ne sont pas elles-mêmes douées de spontanéité mais la manifestent seulement. Troisièmement, ces machines peuvent avoir une réalisation dans le monde de la nature par le biais des ordinateurs, mais cette réalisation est approximative puisque l'idéalité qui gouverne la constitution des machines est indexée sur l'infini des nombres entiers alors qu'en tant qu'expressions, les machines n'apparaissent que comme des textes finis. Quatrièmement, en tant qu'elles sont réalisées dans le fini de la nature, les machines posent deux types de problèmes. Tout d'abord, comme toute autre entité du monde de la nature, on doit déterminer leur *causalité*, en tant que leurs éléments sont produits par la nature d'une part et en tant qu'elles-mêmes produisent des actions d'autre part. Ensuite, on doit préciser leur *finalité*, dans les deux sens de maintien de la cohérence interne du rapport tout/parties d'une part et de tension vers un équilibre dans l'interaction avec le milieu d'autre part.

Ceci posé, nous nous servons d'un texte célèbre de Gödel<sup>68</sup> pour nourrir notre réflexion. Celui-ci établit tout d'abord l'équivalence entre le concept de système formel et celui de procédure mécanique visant la production de théorèmes avant d'en venir aux rapports qu'entretiennent les êtres humains et les machines ainsi qu'aux deux résultats qu'il considère comme ayant été rigoureusement prouvés concernant ce rapport :

«1. L'esprit humain est incapable de formuler (ou de mécaniser) toutes ses intuitions mathématiques. C'est-à-dire : s'il parvient à en formuler certaines, ce fait lui-même produit une nouvelle connaissance intuitive, par exemple la consistance de ce formalisme. Ce fait pourrait être appelé "l'inexhaustivité" des mathématiques. Par ailleurs, en se fondant sur ce qui a été prouvé jusqu'à présent, il demeure possible qu'il puisse exister (et même qu'il soit empiriquement possible de découvrir) une machine à prouver des théorèmes qui soit en fait équivalente à l'intuition mathématique mais dont on ne puisse pas prouver qu'elle l'est ni qu'elle produise seulement des théorèmes corrects de la théorie finitaire des nombres.

---

<sup>68</sup> K. Gödel, (1951), cité dans Hao Wang, ( 1974) : 324.

2. Le second résultat consiste dans la disjonction suivante : ou bien l'esprit humain surpasse toutes les machines (pour être plus précis : il peut décider de plus de questions portant sur la théorie des nombres que n'importe quelle machine) ou bien il existe des questions de théorie des nombres qui sont indécidables pour l'esprit humain.»

Le texte de Gödel fait remarquer la chose suivante : la formalisation de l'intuition mathématique (son expression par le biais d'un algorithme exprimé sous la forme d'une machine de Turing) est seconde par rapport à cette intuition elle-même mais *il peut y avoir*, sans que jamais une preuve puisse en être exhibée, un algorithme qui aurait le même degré de complexité que l'intuition en question.

On pourrait interpréter cette remarque de Gödel en disant qu'il n'y a pas moyen de prouver par le biais d'une procédure mécanique servant de test l'identité entre l'intuition des êtres humains et la machine de Turing spécifique qui actualiserait ce que l'on pourrait appeler le "Grand Algorithme" dont les résultats seraient identiques en tout point à ceux que pourrait obtenir un esprit humain. Mais il me semble que l'on peut aller plus loin dans l'interprétation de ce texte si l'on envisage la "machine", non pas seulement comme une entité *externe* à l'esprit humain mais également comme une forme *interne* à l'esprit, fidèle en cela au fait déjà souligné précédemment que les machines ne sont que l'un des moyens d'*expression* du décalage perpétuel de la pensée avec elle-même. On peut alors se poser la question suivante : quelle est donc, à la source de l'expression humaine, cette structure algorithmique seulement possible (puisque l'on ne peut pas actualiser par une preuve son équivalence avec l'intuition humaine car cette preuve, en servant de nouvel objet à l'intuition, détruirait l'équivalence en question), structure dont l'équivalence avec l'intuition humaine peut cependant être empiriquement retrouvée sous forme de trace partielle (chaque fois que l'on tente de rendre compte par le biais d'un algorithme d'une fonction de l'esprit), mais dont on ne peut jamais être certain qu'elle soit pleinement consistante (puisque il faudrait se situer dans une forme d'intuition supérieure pour en montrer la consistance, forme d'intuition qui nous manque puisque l'on a supposé que ladite structure algorithmique recouvrait toute l'intuition humaine) ?

Il me semble que cette expression d'essence algorithmique, cette "machine" seulement possible, en tant que forme *a priori* d'une expression écrite à la fois spontanée et nécessairement décalée par rapport à l'intuition, est une nouvelle façon d'envisager la notion d'inconscient conçu comme *inconscient procédural*. Il s'agit bien d'un inconscient *procédural* parce que chaque processus mental mis sous forme d'algorithme manifeste la présence d'une générativité

algorithmique de la pensée qui suit la générativité de l'intuition comme son ombre. Mais il s'agit en même temps, et d'une façon paradoxale, d'un *inconscient*, constitutivement "décalé" relativement aux manifestations algorithmiques qu'il engendre. En effet, selon Gödel, l'explicitation de la générativité algorithmique, si elle est partielle, engendre pour l'intuition des occasions de croissance perpétuelle et si elle était totale, deviendrait inaccessible à l'intuition, parce que privée des qualités logiques qui permettraient à l'intuition de lui attribuer un sens. La croissance de l'intuition se trouve donc dépendre d'autre chose que de son fonds propre : elle tire le moteur de son développement de la générativité propre à un potentiel d'écriture systématique, d'essence algorithmique, qui, paradoxalement, ne peut ni ne doit s'explicitier tout entier. Ce potentiel que nous avons imaginé sous la forme seulement possible car imprésentable en totalité d'un Grand Algorithme, se tient, donc, relativement à toute intuition algorithmique particulière, dans la position d'une réserve inconsciente irréductible à toute présentation concrète.

Un point de vue analogue vaudrait également pour l'intelligence artificielle entendue selon le second sens explicité dans l'introduction : le conscient serait alors conçu comme diffusé et constitué dans l'après-coup par le déploiement autonome de ses traces. Qu'une trace puisse valoir dans le temps comme anticipation pour la conscience se comprend mieux dès qu'on la situe dans la réserve active d'un inconscient, extérieur aussi bien qu'intérieur aux sujets psychiques, culturel, technique et matériel à la fois.

Remarquons enfin que Gödel adhère personnellement au premier point de l'alternative qu'il propose touchant l'équivalence entre l'esprit humain et les machines, point selon lequel l'esprit humain surpasse toutes les machines. Pour Gödel, (au moins celui de cette époque puisque la lecture du *Nachlass* semble réserver encore bien des surprises) il serait fondamentalement irrationnel de croire que l'être humain peut poser des problèmes qu'il ne peut pas résoudre. Sans mentionner les raisons que Gödel invoque pour justifier son point de vue et qui tiennent à sa philosophie des mathématiques et sans vouloir trancher le débat, je ferai remarquer que ce point de vue n'est pas contradictoire avec le projet de l'intelligence artificielle consistant à mettre au jour ce que nous avons appelé l'inconscient procédural : toute science, en tant qu'elle vise l'explication des phénomènes, exige de faire l'hypothèse qu'il doit y avoir une explication rationnelle à ce qui apparaît tout d'abord sous la forme d'un problème. Il n'y a pas, de ce point de vue, de véritable alternative car il paraît bien difficile de concevoir une science qui commencerait par considérer que son objet

d'étude est à tout jamais inaccessible, même quand la description de l'objet d'étude en question fait appel à la notion d'inconscient.

#### 4. CONCLUSION

Au terme de ce parcours, je voudrais souligner deux points.

D'une part, en faisant intervenir des données historiques, j'espère avoir montré que la notion d'expérience en intelligence artificielle s'inscrivait dans une longue tradition mais qu'elle mettait l'accent de façon tout à fait spécifique sur un certain nombre de traits caractéristiques de cette tradition. Plus précisément, j'ai essayé de montrer le rôle tout à fait particulier que joue la modélisation informatique dans la transformation de la notion d'expérience puisque la modélisation renouvelle la façon dont on conçoit les rapports de l'objectivité et de l'empiricité.

D'autre part, en essayant d'établir la spécificité épistémologique de la notion d'expérience en intelligence artificielle, j'ai essayé de montrer que la notion d'expérience de pensée y joue un rôle capital. De ce point de vue, la notion de jeu d'écriture ainsi que celle d'inconscient procédural semblent déterminantes dans la mesure où elles permettent d'apporter un nouvel éclairage à ce qu'il faut entendre ici par processus mental : l'effet du transfert à l'intériorité psychique d'une écriture systématique des conduites (humaines ou animales), transfert par lequel l'esprit se découvre comme insondablement calculant.

On peut essayer de donner de ce transfert une multitude d'interprétations philosophiques — par exemple, kantienne, hégélienne ou wittgensteinienne, en faisant varier de surcroît leur tonalité dans un sens plus ou moins favorable à la notion même de transfert. Il est assez facile d'esquisser de telles interprétations même s'il est beaucoup plus difficile de les développer de façon cohérente. Mais l'interprétation qui paraît la moins extrinsèque à son objet serait de dire que nous avons assisté, avec l'intelligence artificielle classique comme avec la psychologie et les philosophies cognitives (mentaliste et comportementaliste) qui lui restent associées, à une forme d'interprétation *mythique* de ces jeux d'écriture, forme qui tend à se présenter comme l'expression d'une *origine* computationnelle de l'esprit dont il tire son identité comme son effectivité et en quoi il peut et doit se reconnaître tout entier.

#### *Bibliographie*

- D. Andler (1990) "Quelle est la place de l'intelligence artificielle dans l'étude de la cognition ?", *Revue internationale de philosophie*, volume 44, n° 172, 1/, Paris, diffusion Presses Universitaires de France.
- D. Andler éd. (1992) *Introduction aux sciences cognitives*, Paris, Gallimard.
- B. Bachimont (1993) "Nature, Culture, Artefacture : la place de l'intelligence artificielle dans les sciences cognitives", *Intellectica*, n° 17, 1993/2.
- C. Bernard (posthume) *Principes de médecine expérimentale*, Paris, Presses Universitaires de France, 1947.
- R. Blanché (1969) *La méthode expérimentale et la philosophie de la physique*, Paris, Colin.
- G. Canguilhem (1965) *La connaissance de la vie*, Paris, Vrin.
- B. Chandrasekaran (1990) "What Kind of Information Processing is Intelligence ?" dans [D. Partridge and Y. Wilks eds., *The Foundations of Artificial Intelligence; a Sourcebook*, Cambridge, Cambridge University Press].
- C. Chevalley (1995) *Pascal, contingence et probabilités*, Paris, Presses Universitaires de France.
- Y. Coppens (1983) *Le singe, l'Afrique et l'homme*, Paris, Fayard.
- L. Couturat (1901) *La logique de Leibniz*, Paris, Félix Alcan.
- M. Davis (1958) *Computability and Unsolvability*, New York, McGraw Hill.
- R. Descartes (1637) *Géométrie*, éd. Adam et Tannery, *Œuvres Complètes*, t.VI, Vrin, Paris, rééd. 1982.
- R. Descartes (1641) *Méditations métaphysiques*.
- R. Dugas (1954) *La mécanique au XVII<sup>ème</sup> siècle*, Neuchâtel, Éditions du Griffon.
- P. Duhem (1908) *SWZEIN TA FAINOMENA, Essai sur la notion de théorie physique de Platon à Galilée*, Paris, Vrin, 1982.
- J.-P. Dupuy (1994) *Aux origines des sciences cognitives*, Paris, La Découverte.
- J. Fodor (1975) *The Language of Thought*, Cambridge, Mass., Harvard University Press.
- K. Gödel (1951) 25th Josiah Willard Gibbs Lecture, cité dans Hao Wang, *From Mathematics to Philosophy*, Londres, Routledge, 1974.
- M. Hallet (1984) *Cantorian Set Theory and Limitation of Size*, Oxford, Clarendon Press.
- Hao Wang (1974) *From Mathematics to Philosophy*, Londres, Routledge.
- J. Hartong (1987) "Le continu et l'ordinateur", *L'ouvert*, 46.
- D. Hilbert (1923) "Les fondements logiques des mathématiques", trad. dans [J. Largeault (1992), *Intuitionisme et théorie de la démonstration*, Paris, Vrin]
- E. Husserl (1935-1936) *La crise des sciences européennes et la phénoménologie transcendantale*, Paris, Gallimard, 1974.
- A. Koyré (1961) *La révolution astronomique : Copernic, Kepler, Borelli*, Paris, Hermann.

- A. Koyré (1971) *Mystiques, spirituels, alchimistes du XVII<sup>ème</sup> siècle allemand*, Paris, Gallimard.
- A. Koyré (1973) “Galilée et l’expérience de Pise : à propos d’une légende” republié dans *Études d’histoire de la pensée scientifique*, Paris, Gallimard.
- D. R. Lachterman (1989) *The Ethics of Geometry; a Genealogy of Modernity*, New-York et Londres, Routledge.
- J. Largeault (1990) Préface au livre de Bernard Ruyer, *Logique*, Paris, Presses Universitaires de France.
- J. Largeault (1992) *Intuitionisme et théorie de la démonstration*, Paris, Vrin.
- D. Marr (1977) “Artificial Intelligence - a Personal View”, *Artificial Intelligence* 9, republié dans [M. Boden éd., *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press, 1990].
- D. Michie (1977) *On Machine Intelligence*, a Halsted Press Book, New York, John Wiley and Sons.
- A. Newell et H. A. Simon (1977) “Computer Science as Empirical Enquiry : Symbols and Search”, The Tenth Turing Lecture, *Communications of the Association for Computing Machinery*, 19, republié dans M. Boden éd., *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press, 1990.
- E. Pacherie éd. (1995) “Fonctionnalismes”, *Intellectica*, n° 21, 1995/2.
- B. Pascal (1645) “Avis nécessaire à ceux qui auront curiosité de voir la machine d’arithmétique et de s’en servir”, *Œuvres Complètes*, éd. L’intégrale, Paris, Le Seuil, 1963.
- B. Pascal (1645) “Lettre dédicatoire à M. le Chancelier sur le sujet de la machine nouvellement inventée par le sieur B.P. pour faire toutes sortes d’opérations d’arithmétique par un mouvement réglé sans plume ni jetons”, *Œuvres Complètes*, éd. L’intégrale, Paris, Le Seuil, 1963.
- B. Pascal (1670) *Pensées*, *Œuvres Complètes*, éd. L’intégrale, Paris, Le Seuil, 1963.
- J. Piaget (1965) *Sagesse et Illusions de la philosophie*, Paris, Presses Universitaires de France.
- R. Penrose (1994) *Shadows of the Mind, a Search for the Missing Science of Consciousness*, Oxford, Oxford University Press.
- J.-M. Salanskis (1991) *L’herméneutique formelle ; L’Infini, le Continu, L’Espace*, Paris, Presses du CNRS.
- J. Searle, (1980) “Minds, Brains and Programs”, *The Behavioral and Brain Sciences*, 3 : 417-424
- D. Sperber (1992) “Les sciences cognitives, les sciences sociales et le matérialisme” dans [D. Andler éd., *Introduction aux sciences cognitives*, Paris, Gallimard].
- R. A. Spitz (1968) *De la naissance à la parole ; la première année de la vie*, Paris, Presses Universitaires de France.



- B. Stiegler (1994, 1996) *La technique et le temps*, I, II, Paris, Galilée.
- R. Thom (1986) “La méthode expérimentale : un mythe des épistémologues (et des savants ?)” dans *La philosophie de sciences aujourd’hui*, Paris, Gauthier-Villars.
- A. M. Turing (1936) ‘On Computable Numbers with an Application to the Entscheidungsproblem’, *Proceedings of the London Mathematical Society*, 42 : 230-265 ; republié dans [M. Davis éd ., *The Undecidable* Raven Press, Hewlett, New-York, 1965 : 115-154].
- A. M. Turing (1950) “Computing Machinery and Intelligence”, *Mind*, vol LIX, n° 236.
- J. van Heijenoort (1967) “Logic as Calculus and Logic as Language”, republié dans [*Selected Essays*, Naples, Bibliopolis, 1985].
- B. Victorri (1995) “Modèles connexionnistes de la mémoire” dans [*Mémoires et amnésies*, Séminaire J.-L. Signoret, F. Eustache et B. Lechevallier, de Boeck, Bruxelles].
- Y.-M. Visetti (1989) “Des systèmes experts aux systèmes de base de connaissance : vers un nouveau schéma régulateur”, *Intellectica* , 1989/1, 7.
- Y.-M. Visetti (1989) compte-rendu du livre de L. Suchman, *Plans and Situated Actions - The Problem of Human/Machine Communication*, *Intellectica*, 1989/1, 7.
- Y.-M. Visetti, (1991)“L’intelligence artificielle est-elle une science ?” dans *L’état des sciences et des techniques*, N. Witkowski dir.
- Y.-M. Visetti (1992)“Intelligence artificielle et systèmes experts : quelques remarques épistémologiques” in *Les nouveaux outils du savoir*, P. Chambat et P. Lévy dirs.
- Y.-M. Visetti (1994) “Les modèles connexionnistes entre perception et sémantique ; note sur la nature et les fonctions des représentations”, *Sémiotiques*, Déc 1994, n° 6-7.
- J. Von Neumann (1966) *Theory of Self-Reproducing Automata*, A. W. Burks éd., Urbana et Londres, University of Illinois Press.
- P. Wagner (1994) *Machine et pensée : l’importance philosophique de l’informatique et de l’intelligence artificielle*, thèse de doctorat de philosophie de l’Université Paris-I.