

**Penser notre relation à la Machine  
(plutôt que nous penser comme des Machines)**

**Denis Berthier**

Institut National des Télécommunications  
Groupe des Ecoles des Télécommunications  
9 rue Charles Fourier, 91011 Evry Cedex, France

[Denis.Berthier@int-evry.fr](mailto:Denis.Berthier@int-evry.fr)

<http://www.carva.org/denis.berthier>

***Introduction***

Penser notre relation à la Machine (ou à son prototype universel, l'Ordinateur) plutôt que nous penser comme des Machines : sous ce titre aux allures de slogan, nous mettons en perspective l'un des thèmes sous-jacents à nos deux livres récents (« Le savoir et l'ordinateur » et « Méditations sur le réel et le virtuel » – respectivement S&O et MRV par la suite). Bien sûr, l'idée générale n'est pas nouvelle en soi, mais, concernant notre relation à la machine, nous tentons de dépasser ce niveau de généralité pour dégager des concepts issus des technologies, mais qui font sens pour les sciences sociales, en adoptant par rapport aux premières ce que nous appellerons la *stratégie du moindre recul nécessaire*.

Cette approche « au ras des techniques » reste aujourd'hui assez rare dans notre domaine d'intérêt, celui des technologies de l'information et de la communication (TIC), peut-être parce qu'elle est très exigeante. Aussi, avant d'entrer dans l'objet de cet article lui-même, ouvrons une brève parenthèse sur notre démarche. Chacun sait quelles sont la diversité, la complexité et l'opacité des objets techniques qui peuplent le monde des TIC. Dans ce monde, notre relation à l'objet technique est problématique, sa polarité

ayant une insidieuse tendance spontanée à s'inverser, en comparaison aux objets classiques : alors que je saisis un marteau, ne suis-je pas plutôt saisi, aspiré, par un monde virtuel ou par un jeu vidéo, ou, pour prendre des exemples un peu moins nets dans la vie quotidienne, par les modalités de la navigation sur Internet ou par la logique d'écriture que m'impose Word® ? Mais le concept d'objet technique est lui-même problématique : l'objet technique, est-ce l'ordinateur ou bien est-ce chaque logiciel ; et l'ordinateur n'est-il dans ce dernier cas que le support matériel de cette infinité potentielle d'objets, dont la potentialité s'actualise en outre à une vitesse qui dépasse l'entendement du profane – processus par lequel l'universalité théorique de l'ordinateur (en tant que machine de Turing) se déploie concrètement dans l'historicité ? Pire, dans le cas d'Internet, qu'est-ce que cet objet informe, distribué sur toute la planète ? D'où la difficulté de l'épistémologie à appréhender les objets techniques du monde des TIC, tandis que, de leur côté, les sciences sociales ne les approchent que de l'« extérieur », en termes d'usages, d'impact social, de relations à l'imaginaire collectif, etc.

C'est pourquoi, pour être disponibles et exploitables par les sciences sociales au moment utile, compte tenu du temps nécessaire à leur élaboration, les concepts que nous visons à dégager doivent être formulés en termes susceptibles de s'appliquer encore aux technologies qui apparaîtront en toute probabilité à court terme (en évitant de tomber dans l'anticipation et la prospective incontrôlées). Notre approche ne peut donc se contenter d'un regard extérieur sur des objets (logiciels) existants, comme le fait le sociologue ; les concepts qu'elle recherche ne doivent pas être liés à une application particulière, ils doivent viser le même type d'universalité que l'ordinateur. Nous devons donc remonter aux principes fondamentaux pour saisir ce qui, dans leur généralité, fait sens en dehors de leur univers technique.

Venons-en au contenu de cet article. Sous des formes multiples, le programme cybernétique, « nous penser comme des Machines » (section 1), imprègne toujours, souvent de manière souterraine, les fameuses « sciences cognitives » – ou, plutôt, certains discours tenus dessus, sur la base d'une confusion instaurée par Wiener (section 2). Ensemble, les deux premières sections explicitent les enjeux culturels de notre questionnement.

Face à la situation décrite, nous considérons (section 3) qu'il est nécessaire de revenir sur certaines des technologies qui ont donné corps à ce programme, il y a déjà plus d'un demi-siècle, pour rappeler, sur la base de leur réalité actuelle, le programme alternatif moins réducteur qu'elles imposent en fait : « penser notre relation à la machine ». Nous appuyons ces considérations sur le cas de l'intelligence artificielle (IA). Nous évoquons en particulier (section 4) les raisons culturelles pour lesquelles l'IA et les types de questionnements qu'elle soulève ne sont pas près d'être évacués de notre univers sémiotico-cognitif – justifiant qu'on doive en permanence remettre sur le métier ce programme alternatif.

Elaborer notre relation à la machine conduit naturellement (section 5) à des rapprochements conceptuels inattendus entre l'IA et une autre des disciplines phares des nouvelles technologies, la réalité virtuelle (RV). Apparaît dès lors la nécessité de poser plus généralement le virtuel comme une catégorie épistémologique majeure (ou une catégorie ontologique, pour ceux que ce mot ne dérange pas), clairement distinguée de l'imaginaire et du potentiel, avec lesquelles elle est trop souvent confondue. Une telle démarche va à l'encontre de tous les usages abusifs auxquels le terme de virtuel donne actuellement lieu – usages qui contribuent à le vider de tout sens. Or, cette perte de sens n'est pas innocente, car, en rendant la technique seule responsable de tous nos maux, elle sert à camoufler des problèmes qui sont avant tout des problèmes de vie en société (comme un déficit de communication effective, en dépit de l'explosion des technologies du même nom, etc.). De son fait, de nombreux discours qui se réclament du virtuel contribuent

involontairement à entretenir ses mythes (société virtuelle, etc.), alors même qu'ils veulent parfois les dénoncer.

### ***1. Nous penser comme des machines : de la cybernétique aux sciences cognitives***

Vouloir égaler les dieux en créant un homme, ou un simulacre « à l'image de l'homme » (selon le titre d'un livre de Philippe Breton), n'est pas une idée neuve, surtout si l'on comptabilise les procédés plus ou moins magiques, comme d'animer une statue en « soufflant sur la glaise ». A partir de l'époque classique, cette prétention abstraite a connu un premier changement de nature, en même temps qu'elle se spécialisait, lorsqu'on a conçu non plus l'homme, mais seulement certains de ses comportements, comme un mécanisme, et/ou qu'on les a reproduits avec une *machine*. A divers titres, les noms de Descartes, Pascal, Leibniz, Vaucanson, Babbage et lady Ada Lovelace s'imposent à l'esprit.

Au milieu du XXe siècle, cependant, ces prétentions ont pris une tout autre tournure et une tout autre dimension avec l'apparition de la cybernétique de Norbert Wiener. A l'origine purement technique, celle-ci, sous l'impulsion de Wiener lui-même, a en effet rapidement servi de caution scientifique à ce qui est devenu une véritable idéologie, consistant à concevoir l'homme comme une sorte d'automate et la pensée comme un calcul. Cette idéologie, qui s'adosse désormais aussi aux fantasmes qu'inspirent la science-fiction et les robots, imprègne encore les conceptions populaires de l'esprit, de la pensée, de la mémoire, etc., et continue à être largement soutenue (souvent d'une manière implicite très insidieuse), en général pas au sein même des disciplines scientifiques formant la nébuleuse des « sciences cognitives », mais par les médias traitant de ces sujets (dans lesquels nous incluons, dans tout cet article, certains prophètes des nouvelles technologies et certaine sociologie à tendance journalistique).

A cet égard, nombreux sont les auteurs, à commencer par Steve Heims (1991), Philippe Breton (1992) et Jean-Pierre Dupuy (1994), à avoir explicité les filiations entre les récentes sciences cognitives et la défunte cybernétique, ainsi que les lourdes conséquences intellectuelles, culturelles et sociétales de certaines de leurs interprétations et vulgarisations conduisant à penser l'homme comme une machine. Ces dérives, auxquelles échappent certes la majorité des chercheurs dans ces disciplines, restent d'autant plus difficiles à contrer en dehors des milieux spécialisés, que les médias ont tendance à se focaliser sur le sensationnel, qui est souvent le point de vue le plus marginal. Je vois ainsi chaque année des dizaines d'étudiants en écoles d'ingénieur ou de management, qui, avant tout cours sur le sujet, ne connaissant l'IA qu'à travers le film de Spielberg (ou ses équivalents), pensent que l'IA a pour but de reproduire l'intelligence humaine.

La portée culturelle de ces dérives est directement proportionnelle au fait que la machine à laquelle nous sommes assimilés se présente sous des formes extrêmement variées, ayant chacune leur constellation conceptuelle spécifique et parlant chacune différemment à l'imagination (et dont il est peu pertinent ici que certaines soient formellement équivalentes à d'autres) :

1) dans la « première » cybernétique (Wiener), un système doté de complexité comportementale grâce à ses boucles de rétroaction ; la notion de *rétroaction* (liée à celle de complexité), extrêmement novatrice à l'époque quoique scientifiquement très élémentaire, a fait l'objet d'un véritable culte, initié par Wiener lui-même et qui s'est largement étendu jusque dans les sciences sociales (Morin) ;

2) dans la « deuxième » cybernétique (Von Foerster, Atlan, Maturana, Varela), un système auto-organisateur, auto-adaptatif, auto-poïétique ; l'idée centrale est le principe d'« *ordre par le bruit* » (selon lequel un système tire son organisation par évolution sous l'effet du « bruit » de son environnement) ; aujourd'hui, les idées du chimiste Prigogine sont souvent associées à celles de cette école ;

3) dans les approches « neuronales » (McCulloch, Pitts), un réseau de neurones formels, accomplissant un calcul massivement parallèle ; la fascination initialement produite par ce type de machine « analogique » (c'est-à-dire continue) provient du fait que les opérations discrètes de la logique formelle peuvent *émerger* de son fonctionnement ; aujourd'hui, la théorie des catastrophes de René Thom est souvent appelée à la rescousse pour expliquer cette *émergence* ;

4) dans les approches d'origine purement informatique (Turing, Von Neumann), une machine de Turing universelle (un ordinateur) ; c'est l'*universalité*, c'est-à-dire la capacité à reproduire tous les comportements (assimilés à toutes les fonctions d'entrée-sortie calculables), qui est ici considérée comme essentielle ;

5) plus spécifiquement, en IA « symbolique » (Newell), un « système de symboles physiques » (ou système de signes) ; ici, la fascination initiale tient à la faculté de ces systèmes à résoudre, par des *manipulations formelles de symboles a-signifiants*, des problèmes censés demander de l'intelligence (preuve de théorèmes mathématiques, jeux d'échecs, etc.). Nous nous attarderons plus loin sur ce cas, mais notons dès à présent que cette IA « symbolique » devrait tenir une place très particulière dans nos réflexions, puisque nous interagissons « naturellement » entre nous au moyen de signes et non directement entre nos cerveaux ou nos neurones.

Il reste, *in fine*, deux affirmations complémentaires, qui constituent la forme la plus extrême sous laquelle le cognitivisme peut se retrouver présenté : le cerveau est une machine, la pensée est le calcul effectué par cette machine. Dans sa version « connexionniste » ou « neuronale », la machine est un réseau de neurones formels ; dans sa version « symbolique », la machine est plus abstraite, c'est un « système de représentation et traitement de connaissances » (par exemple un « moteur d'inférences » qui va

faire des déductions à partir d'expressions logiques représentant des connaissances, ou bien un « réseau sémantique », qui va pouvoir établir des associations entre idées ou concepts).

Si, à l'origine, la machine a pu être conçue « à l'image de l'homme », Dupuy (1994, p. 41) note que : « ce n'est pas la machine que les premiers cybernéticiens dotent d'humanité – c'est l'humain que, délibérément, ils assimilent à la machine ». Cette affirmation trouve des illustrations dans les titres ou sous-titres évocateurs (à intention critique, pour certains) d'un nombre incalculable de livres et d'articles. Pour fixer les idées, donnons un petit échantillon de livres : « Machines à penser » (Arsac) ; « Thinking machines » (Pratt) ; « Le cerveau-machine » (Jeannerod) ; « L'âme machine » (Ganascia) ; « De la machine à penser à la pensée machine » (Lafontaine).

## ***2. Les discours identificateurs ; source et parade***

Rien, dans la pratique scientifique normale de la modélisation, ne justifie les identifications hâtives de l'homme à la machine. Le fait qu'on ait développé un modèle plus ou moins adéquat de certains aspects, même complexes, d'un « objet » (ici le cerveau ou la pensée) n'autorise pas à prétendre que l'objet est identique au modèle. Evidemment, cela autorise encore moins à faire du modèle (au sens scientifique du terme) une référence à laquelle l'objet devrait normalement se conformer, qu'il devrait « prendre comme modèle », par un jeu pernicieux sur l'ambiguïté du terme dans le langage courant (comme le souligne Dupuy 1994).

Ces remarques, aussi évidentes et peu originales soient-elles, n'ont pourtant rien de superflu, puisque l'ambition explicite de Wiener, et de bien d'autres après lui, était justement de réformer la société en donnant le contrôle aux machines, qu'il considérait plus « parfaites » que l'homme. Il serait dangereux de croire qu'elles seraient périmées, car un discours en tous points analogue à celui de Wiener se retrouve aujourd'hui chez certains promoteurs des « super-intelligences » (Kurzweil, etc.) ou des nanotechnologies.

Malheureusement, toute l'aventure des discours tenus sur la cybernétique et les sciences cognitives est entachée de grossières erreurs épistémologiques de cette nature, souvent propagées par une médiatisation outrancière. Avec le risque majeur, en ce qui concerne l'homme, qu'en tant qu'objet de modélisation, il soit tellement contraint par les images de lui que ces modèles lui renvoient sans cesse, et souvent de manière très occulte, qu'il finisse par se considérer comme véritablement semblable au modèle. Par exemple, l'idée, issue en droite ligne de l'omniprésente informatique, que la mémoire est un ensemble de cases à remplir n'est-elle pas tellement communément admise que la pratique « standard » de la pédagogie, fondamentalement basée sur le gavage (donner les réponses avant même que l'élève ait eu le loisir de se poser les questions), ne provoque guère de réactions ?

Nombreux sont les auteurs à avoir désigné Wiener (avec aussi parfois McCullough – voir Dupuy 1994) comme l'initiateur moderne de cette manière de bâtir une vision générale du monde à partir de faits scientifiques ou technologiques. Un peu moins nombreux sont ceux à avoir dénoncé clairement l'illégitimité du procédé. Illégitimité pourtant flagrante, si l'on considère le fossé conceptuel qui sépare les deux ouvrages fondateurs de Wiener : l'un purement technique (« Cybernetics »), le deuxième très largement utopique ou idéologique (« The Human Use of Human Beings », pudiquement traduit en français comme « Cybernétique et société »).

Aucune science ou technique ne peut *légitimer* aucune philosophie ; tout au plus pourrait-elle la contredire. Ainsi, quand des auteurs très médiatisés affirment que la cybernétique ou l'IA ou les réseaux de neurones formels conduisent à la déconstruction du Moi ou du Sujet, thèmes favoris du « déconstructionnisme », (d'où l'on conclut par exemple que le bouddhisme et la cybernétique, c'est « kif-kif » – Lafontaine 2004), le fossé conceptuel entre les termes me semble si énorme que j'avoue ne

strictement rien comprendre, quelle que soit la masse de considérations intelligentes dont on veuille tenter de le combler ou le masquer. Cependant, il ne suffit pas de dénoncer une erreur, surtout lorsqu'elle est aussi profondément ancrée dans certains esprits, pour y remédier durablement.

Il ne suffit pas non plus, à notre avis, de détourner la question en démontrant (comme le font en détail Breton et Lafontaine) les liens étroits que les disciplines concernées (et leurs représentants) ont, depuis leurs débuts, entretenus avec le complexe militaro-industriel, ou de souligner le rôle effectif qu'elles ont joué pendant et après la guerre ; car le risque serait alors grand d'induire une condamnation massive de ces disciplines et d'alimenter la technophobie ambiante, laquelle dispensera à coup sûr d'analyser leur impact réel. Sur ce point, nous rejoignons Dupuy (1994) : « Je crois trop au pouvoir des idées et à leur dynamique autonome pour me satisfaire en général de la perspective "externaliste" en philosophie des sciences ». Ces propos renvoient à une distinction nette entre épistémologie et sociologie des sciences. Chacune garde évidemment son intérêt propre, d'autant que, pour tout compliquer, les discours d'accompagnement des sciences et des techniques ont parfois, au moins pour un temps, un impact socioculturel plus important que les sciences et techniques elles-mêmes, et souvent d'autant plus important qu'ils les dénaturent. Un exemple flagrant se trouve dans Breton (1992 ou 1997) et Lafontaine (2004), qui, en fonction des discours dont elle s'entoure, ont cru pouvoir inscrire l'IA « symbolique » dans le courant de la « philosophie cybernétique ». Mais, jamais ces auteurs ne font référence ni à Newell, ni à son fameux article de 1980, où il explicite la conception que se fait l'IA de ce qu'est un système de signes, ni à aucun texte qui serait vraiment spécifique à l'IA – bref, à rien de ce qui permet d'affirmer qu'il n'y a aucune raison de figer l'IA dans une telle interprétation.

La section 3 a pour but de rappeler qu'on obtient une vision de l'IA fort différente dès lors qu'on prend la peine d'explicitier la nature effective de notre relation aux artefacts qu'elle produit, en

l'occurrence les « agents intelligents » (expression qu'il faut comprendre comme un terme technique indécomposable, sans se focaliser sur le mot « intelligent »). Le cas de cette « bonne vieille » IA est d'autant plus intéressant, comme il sera montré à la section 4, que ses principes de base trouvent des résonances au plus profond de notre culture et que ses applications semblent devoir continuer inéluctablement leur déploiement massif (contrairement à feu la cybernétique). En outre, ce n'est qu'en explicitant cette relation véritable que nous avons quelque chance d'en rester maîtres. Conception certes subversive pour tous ceux qui ont intérêt à nous endormir en propageant l'idée fataliste que notre avenir est régi par des lois (formulées en termes de marché, de mondialisation, de société de l'information, etc.) sur lesquelles nous n'avons aucune possibilité d'influer. La section 5 montrera qu'une démarche identique doit être accomplie autour de la notion de « virtuel ».

### ***3. Penser notre relation à la Machine (1) : l'interopérabilité sémiotico-cognitive***

Commençons par repérer le fossé intellectuel qui sépare l'IA de la cybernétique, avec les deux sauts conceptuels successifs que suppose le passage de l'une à l'autre. Ceux-ci concernent la notion centrale d'information. Tout d'abord, à la conception de Wiener, qui lui donne une dimension physique et l'assimile à la néguentropie, quantité *continue*, s'oppose la conception communicationnelle beaucoup plus abstraite de Shannon, qui sera aussi celle de l'informatique, laquelle en fait une quantité *discrète*, exprimable en « bits ».

Quant au deuxième saut, partons de la remarque triviale que tout artefact (au sens large – incluant un programme informatique) doit impérativement être pensé selon plusieurs perspectives. Au minimum, un artefact est d'une part un objet physique, soumis

comme tel à la *causalité* physique, et d'autre part un objet conçu pour remplir certaines fonctions, répondant comme tel à une certaine *finalité* – deux points de vue qui ont toujours été considérés par la philosophie classique comme incompatibles et qu'il faut bien pourtant admettre simultanément. C'est là un impact culturel majeur des technologies, « nouvelles » ou pas, qu'elles remettent en cause certains préjugés philosophiques anciens.

Ainsi l'IA suppose-t-elle un changement de regard sur l'« information » : le matériau de l'IA, la « connaissance », exige un point de vue conceptuel beaucoup plus élevé. Un système d'IA ne fait pas des manipulations arbitraires de suites de bits indifférenciés ; il traite de la « connaissance » au sens technique précis où Word® traite du « texte » et Photoshop® traite des « images » – c'est-à-dire qu'il effectue des opérations très spécialisées sur des objets fortement structurés, de façon que le tout puisse être interprété, par nous humains, en termes de déductions logiques, d'associations de concepts, etc. Il ne s'agit évidemment pas ici de nier la réalité ni des opérations « logiques », au niveau des bits informationnels, ni des processus physiques sous-jacents, au niveau des électrons, mais il s'agit de reconnaître qu'un tel système *doit* impérativement être décrit à plusieurs niveaux conceptuels différents. En outre, au-dessus de ces niveaux techniques, Newell (1982) a aussi introduit un « niveau des savoirs », où la connaissance est formulée de manière indépendante de toute représentation informatique.

Toutes les méthodologies de développement en IA partent de ces idées classiques, sans lesquelles elles n'auraient aucun sens, sans lesquelles aucun projet d'IA ne pourrait voir le jour. (Après coup, on se dit évidemment que ces remarques auraient dû être appliquées à la cybernétique – mais elles ne le furent pas.) Ce n'est donc pas par hasard que, dans S&O, nous avons attaché beaucoup d'importance aux méthodologies (alors que, dans les réflexions philosophiques sur l'IA – surtout celles où il est question de l'intelligence, de la cognition, de l'esprit, etc. – elles sont généralement passées sous silence, reléguées dans la catégorie des

problèmes d'ingénieurs, peu dignes d'intérêt). Bien loin des discours philosophiques usuels sur l'IA, elles expriment la réalité de ce secteur d'activité ; c'est à travers elles que se révèle le plus crûment le lien véritable entre l'artefact, son concepteur et son utilisateur. (Cela ne semble pas devoir être spécifique à l'IA. En physique, la vision épistémologique de Mme Mugur-Schächter (2002) repose fondamentalement sur l'analyse des protocoles expérimentaux de la mécanique quantique).

Que déduire aujourd'hui de ces méthodologies en IA ?

1°) L'IA a pour objet de construire des « agents artificiels » dotés de « buts » et de « connaissances » et régis par un principe de rationalité simple : utiliser rationnellement (ce qui laisse pas mal d'indétermination) ses connaissances pour atteindre ses buts.

2°) Les connaissances en question peuvent être exprimées dans des systèmes formels de représentation, sous une forme à la fois compréhensible par un humain ne sachant rien des subtilités informatiques d'une machine capable de les mettre en œuvre et exploitable par une machine ne « sachant » rien des subtilités psychiques d'un humain capable de les comprendre.

3°) Ces agents artificiels sont capables de communiquer entre eux à l'aide de primitives de communication formalisées, inspirées de la théorie des actes de parole d'Austin et Searle ; au-dessus de ces capacités communicationnelles « primitives », ils peuvent aussi être soumis à divers modes d'interaction « sociale ».

4°) Avec les avancées de la linguistique informatique, ces agents artificiels sont aussi capables de communiquer avec nous, en un langage de plus en plus proche de notre langage naturel ; en outre, la communication entre eux et nous commence à exploiter diverses modalités extra-linguistiques (schémas, expressions faciales, etc.).

5°) Avec l'IA et la linguistique réunies, l'informatique passe ainsi du traitement de l'information au traitement des savoirs et permet un partage de savoirs formalisés entre l'homme et l'ordinateur – au

sens technique précis où tous deux peuvent communiquer et raisonner à leur sujet.

Dans MRV, nous avons réinterprété ces résultats. *L'objectif général de l'IA peut en effet se reformuler en termes d'interopérabilité : mettre en place une interopérabilité sémiotico-cognitive de plus en plus étendue entre l'homme et l'ordinateur.*

Rappelons que le concept purement technique d'*interopérabilité* (litt. la capacité à opérer, ou à fonctionner, ensemble) tient aujourd'hui une place centrale dans les technologies de l'information et des réseaux. Dans S&O (chapitre 9 et partie 4) et dans MRV (chapitre 6), nous avons pu l'étendre de manière naturelle en celui d'une interopérabilité sémiotico-cognitive entre l'homme et l'ordinateur. Notons déjà que ce terme donne à réfléchir : car l'interopérabilité n'est pas l'identité ; en pratique, c'est même quasiment le contraire, la question ne se posant que pour pallier le défaut d'identité (des machines, des systèmes d'exploitation, des langages de programmation, etc.). Il s'oppose donc à tous les discours identificatoires. En outre, il véhicule une plus grande exigence que le terme courant d'interface (lequel semble ne viser que la surface des choses). Il en résulte qu'*il ne devrait plus être possible de faire accroire que l'intelligence artificielle nous invite à nous penser comme des Machines ; car, à travers les développements et les avatars successifs de cette interopérabilité, elle nous presse, au contraire, de repenser en permanence notre relation à la Machine.* Faute de quoi l'apparition de chaque nouvelle technologie (simulation de traits de caractères, de sentiments, d'expressions faciales, synthèse vocale réaliste, reconnaissance des gestes, etc.) sera inévitablement interprétée par les médias en termes identificatoires.

#### ***4. Penser notre relation à la Machine (2) : ordre symbolique, structuralisme, dimension culturelle de l'IA***

A l'opposé des discours rémanents sur l'échec de l'IA « symbolique » (discours issus de la grande désillusion des années

90 et généralement liés à ses interprétations identificatoires abusives), il nous faut partir du constat de ses développements effectifs et mesurer l'ampleur du champ des savoirs effectivement atteint en application de ses principes. Quant à l'avenir proche, on remarquera qu'un accès efficace aux masses de connaissances que recèlent les milliards de « pages » du Web ne pourra se faire que de manière simultanée pour l'homme et l'ordinateur, à l'aide d'« agents d'information intelligents » ayant accès à une partie significative du sens des contenus – objectif auquel répondent en partie les technologies récentes du « Web sémantique » ou du « Web intelligent ». Sur ces points, faute de place, nous ne pouvons que renvoyer à S&O (chapitre 10).

Devant ces succès réels de l'IA (sans lesquels nos réflexions n'auraient pas d'intérêt), succès trop souvent méconnus hors des cercles spécialisés, une question s'impose cependant : comment tout cela est-il possible ? Une première réponse, selon une démarche classique et purement technique, montrerait qu'une machine de Turing (un ordinateur) a la capacité à représenter des connaissances et à faire dessus certains types de traitements (déductifs, associatifs, etc.) ; et l'on achèverait alors la démonstration par l'existence *de facto* de gigantesques bases de connaissances : d'expertise, de sens commun, ou linguistiques.

Mais il est indispensable de compléter ces propos par un autre type de réponse, non technique, qui, en établissant ses liens conceptuels profonds avec le structuralisme, permet d'élucider pleinement *la dimension intrinsèquement culturelle de l'IA* ; d'elle dépendent fondamentalement toutes les prévisions que l'on peut faire sur sa portée et ses impacts à venir. Cette deuxième réponse (que nous avons développée dans MRV, chapitre sept) s'entend sur deux plans distincts mais indissociables :

- au niveau des systèmes de signes, l'IA et le structuralisme (dont il suffit ici de retenir la composante scientifique, ses multiples interprétations, extensions ou dérives philosophiques n'étant pas

utiles à notre propos) ont des conceptions extrêmement proches de ce qu'est un système de signes, toutes deux étant basées sur l'idée d'une manipulation formelle de « signes » ou « symboles » a-signifiants (Berthier 1994). Sans doute le structuralisme a-t-il été sur ce point influencé dans certaines de ses formulations par la cybernétique (Breton 1992, Dupuy 1994), mais l'essentiel n'est pas là. Car de nombreux faits (linguistiques, anthropologiques, etc.) étayaient directement un tel modèle de l'« ordre symbolique ».

- au niveau des savoirs, la pratique actuelle de la modélisation, base de toute l'activité scientifique, reste de nature profondément structuraliste, au moins dans ses phases structurantes initiales (recherche de différences significatives et d'homologies structurales). Ainsi, seules comptent des relations entre éléments n'ayant pas de sens intrinsèque, pas d'essence, etc. – condition nécessaire pour que des savoirs les concernant puissent être formulés dans un système de signes formels. On songera aussi bien aux phonèmes, n'ayant d'existence qu'en vertu de l'ensemble des jeux d'oppositions phonématiques, qu'aux particules élémentaires de la physique, auxquelles on ne sait attribuer de « nature propre » en dehors des « groupes d'invariance » qui les relie. En ce sens, le structuralisme, comme l'IA, court-circuite la question ontologique (au sens philosophique de ce terme, pas au sens technique qu'il a en IA).

Nous admettons facilement avec Dupuy (1994, p. 173) et d'autres que certaines formulations structuralistes ont terriblement vieilli. C'est cependant le lieu de rappeler qu'un modèle est toujours par nature un point de vue *partiel* sur son objet, que notre faculté symbolique ne doit donc pas être totalement assimilée au modèle formalisé qu'en offrent le structuralisme ou les systèmes de symboles physiques de l'IA ; ce qui laisse place à l'étude (par la philosophie, la poésie, etc.) d'un éventuel « reste » non formalisable (ou non calculable) – reste qui pourrait s'avérer *in fine* l'essentiel du point de vue d'un Sujet. Quels que soient les territoires gagnés par le formalisé sur le « reste » (et l'IA « symbolique » démontre leur ampleur), quels que soient les

nouveaux « restes » produits en retour (car il n'y a aucune raison de penser que le « reste » est prédéfini), il demeure qu'en vertu des remarques précédentes, avec l'IA et ses « systèmes de symboles physiques », l'insistance de cet ordre symbolique formel apparaît comme l'un des faits les plus incontournables de ce début du 3<sup>ème</sup> millénaire.

En raison de la place centrale du structuralisme dans la culture du XX<sup>e</sup> siècle, en raison de ses liens conceptuels étroits avec l'IA, et plus seulement (comme à la section 3) en raison d'une dynamique propre des sciences et techniques, les « agents intelligents » occuperont une place grandissante dans notre univers sémiotico-cognitif et l'IA « symbolique » nous imposera une certaine relation sémiotico-cognitive formelle à l'ordinateur – relation en perpétuelle évolution, motivant une réouverture permanente de nos interrogations sur le sens, sur le savoir, sur notre relation aux systèmes de signes formels et aux agents artificiels qui les implémentent, ainsi que sur ce qui continue à nous distinguer d'eux. *Et c'est aussi parce qu'elle a ces liens conceptuels étroits avec le structuralisme qu'il est légitime pour l'IA de viser à l'interopérabilité sémiotico-cognitive.*

### ***5. Penser notre relation à la Machine (3) : l'interopérabilité sensori-motrice et la catégorie épistémologique du virtuel***

Il se trouve que, si l'IA a pour but d'établir une interopérabilité entre l'homme et l'ordinateur dans l'ordre sémiotico-cognitif, la réalité virtuelle (RV) s'assigne un but analogue dans l'ordre sensori-moteur. Sans entrer dans les détails (qu'on trouvera dans MRV, chapitre 5), cette nouvelle définition de la RV n'est qu'une reformulation de sa définition habituelle en termes d'immersion sensori-motrice dans des mondes virtuels. *IA et RV visent ainsi conjointement à établir l'interopérabilité dans les deux ordres majeurs de l'expérience humaine.*

Or, si le concept d'interopérabilité peut assez facilement être exporté de l'informatique générale ou de l'IA vers la RV, celui de virtuel peut être exporté en retour de la RV vers l'IA, avec un peu plus de travail. A cette fin, en nous appuyant sur une analyse de la racine latine (*virtus*), et en nous basant aussi sur des usages du terme bien établis dans des cas spécifiques (mondes virtuels de la réalité virtuelle, images virtuelles de l'optique géométrique, particules virtuelles de la physique, etc.), nous avons commencé dans MRV par définir le virtuel de manière radicalement nouvelle comme une catégorie épistémologique générale : *le virtuel est ce qui, sans être réel, a, avec force et de manière pleinement actuelle, les qualités (propriétés) du réel*. Cette catégorie se trouve ainsi distinguée clairement de celles de l'imaginaire et du potentiel, avec lesquelles elle est généralement confondue.

Par exemple, il n'y a rien d'imaginaire ou de potentiel dans mon reflet dans un miroir (qui est, techniquement parlant, une image virtuelle). Bien que non réel, mon reflet n'est pas en attente d'une quelconque actualisation ; il est pleinement actuel ; en particulier, les rayons lumineux « issus de lui » ne changeront pas leur trajectoire, que je sois là ou non pour l'observer. C'est sur lui (et non sur le plan du miroir) qu'un appareil photo autofocus (tout comme l'œil) fera le point, que je sois là ou pas. Je conseille vivement de faire l'expérience, car nombre de personnes interrogées à ce sujet (et surtout parmi les non scientifiques) pensent que le point se fera sur le miroir, parce que c'est le seul objet ayant une réalité physique. Et quand ils constatent qu'il se fait sur l'image virtuelle, ils ne comprennent pas qu'un appareil puisse « voir » un objet qui « n'existe pas ». L'objection n'est d'ailleurs pas du tout stupide, car la magie apparente du résultat dépend étroitement de l'algorithme mis en œuvre par les systèmes de mise au point, qui est basé sur les micro contrastes. S'il reposait par exemple sur des techniques de télémétrie laser, ça ne marcherait pas. Ce qui montre que, bien que pleinement actuel, le virtuel n'a d'existence (il n'ek-siste, littéralement) que pour un système perceptif déterminé (et la même remarque s'applique aux

mondes virtuels de la réalité virtuelle, qui sont conçus en fonction de nos capacités perceptives à nous humains).

Il faut noter que notre conception du virtuel se distingue, par exemple, de celle de Pierre Lévy (chantre de l'Internet et du virtuel) qui, à la suite de Deleuze (1968), oppose le virtuel à l'actuel (et pour qui, en conséquence, le virtuel et le potentiel ne sont pas clairement différenciés) ; mais, avec cette opposition, la toute puissance du virtuel à nous leurrer reste inexplicée. Notre conception se distingue aussi de celles qui voudraient assimiler le virtuel au numérique ; car ce serait en restreindre considérablement le champ, d'une manière non justifiée par l'étymologie et qui le soumettrait arbitrairement, tout en l'alimentant, au mythe de la société de l'information.

Une conséquence quasi tautologique de notre définition est qu'il est impossible de distinguer le réel du virtuel *par des principes généraux* (bien que, dans chaque situation particulière, cela ne présente habituellement pas de difficulté pour une personne normale). Quoique sous une forme très différente, cela répond à l'un des thèmes majeurs des films sur la RV (Matrix, eXistenZ, etc.) : la difficulté à établir la distinction. Que celle-ci échappe à la formalisation n'empêche pas qu'il soit nécessaire d'en maintenir le principe (ne serait-ce que pour échapper au solipsisme), de même que nous avons maintenu une distinction de principe entre l'ordre symbolique formel et notre faculté symbolique. De la même manière aussi, notre incapacité à établir des *preuves du réel* pointe vers la nécessité de développer notre *sens du réel*.

On peut alors appliquer cette définition aux agents artificiels de l'IA et constater que, sans être des agents réels, ils ont, dans l'ordre sémiotico-cognitif, toutes les propriétés et comportements habituellement attendus d'un agent : ce sont donc des agents virtuels. De même, tous les états mentaux et catégories cognitives sous lesquels on a tendance à les percevoir (buts, intentions,

intelligence, etc.) doivent être considérés comme virtuels. Ce vocabulaire permet d'évacuer toutes les fausses questions soulevées à propos des sciences cognitives : un agent artificiel est-il intelligent, a-t-il une conscience, etc. ? Dans notre approche, tous ces prédicats peuvent éventuellement lui être attribués, mais seulement comme virtuels. Il est en outre impossible de se prononcer sur leur statut de réalité. Ainsi, les affirmations identificatoires de l'IA « forte » se retrouvent-elles purement et simplement dissoutes. *Le but de l'IA est d'établir une interopérabilité sémiotico-cognitive avec des agents virtuels, de même que le but de la RV est d'établir une interopérabilité sensorimotrice avec des mondes virtuels.*

En définissant ainsi le virtuel, nous confortons le sens que nous avons tendance à attribuer spontanément à un certain type de perceptions ou d'interactions. Comme pour mon reflet dans un miroir, on parlera pour la RV d'immersion dans un monde virtuel plutôt que d'illusions perceptives. De même, pour l'IA, on parlera d'interactions sémiotico-cognitives avec des agents virtuels plutôt que de l'illusion d'un dialogue avec un ordinateur. Un objet technique, avons-nous dit, peut être décrit à plusieurs niveaux d'abstraction. Pour un agent artificiel, ce peut être comme la circulation d'électrons dans un circuit, mais cela a peu d'intérêt ; ce peut être comme un certain programme logique, et cela commence à avoir un intérêt pour l'informaticien ; mais ce peut aussi être comme un agent virtuel capable de certaines interactions sémiotico-cognitives et cela a un intérêt pratique immédiat pour tous ses interlocuteurs humains – exactement comme la notion de reflet a un intérêt évident du point de vue phénoménologique. Ces exemples montrent combien la notion d'interopérabilité pose sur les systèmes informatiques des conditions plus profondes que celle d'interface et combien elle est étroitement liée à celle de virtuel : il n'y a interopérabilité sensorimotrice que si j'ai l'impression d'être dans un monde (qui est, de fait, virtuel), il n'y a interopérabilité

sémiotico-cognitive que si j'ai l'impression d'interagir avec un agent (qui est, de fait, virtuel).

Mais, dira-t-on, ce vocabulaire n'a-t-il pas d'effets pervers ? Car où s'arrêter ? Dois-je aussi parler de l'âme ou de la conscience de mon agent virtuel ? Réponse : rien dans nos interactions avec un agent virtuel (pas plus qu'avec un marteau) ne permet (ni ne nécessite) de dire qu'il a une âme ou une conscience, même virtuelle (sauf au cinéma, avec, par exemple, le personnage principal du film « Passé virtuel »). Et pour son intelligence ? Il nous apparaît bien comme ayant une certaine forme d'intelligence (mais alors assez limitée, au moins en l'état des technologies, car il lui manque justement la généralité). Et pour ses buts ? Il a bel et bien des buts (virtuels), puisque ses « buts » informatiques opèrent, du point de vue extérieur de l'interlocuteur humain, comme opèrent des buts réels.

Mais alors, dira-t-on encore, qu'en est-il de moi-même ? Comment puis-je savoir si j'ai bien un Moi réel ? Réponse : en aucune manière. Pour Francisco Varela (2004), le Moi est virtuel – ce qu'il présente comme une interprétation moderne de la doctrine Mādhyamika de l'*anatman* (ou Non-Moi) ; Varela utilise le terme « virtuel » de manière informelle, mais parfaitement compatible avec notre définition. Le moi est virtuel : notre définition du virtuel permet en outre de donner un sens plus fort que celui d'une simple métaphore aux divers schémas optiques qui parsèment le Séminaire I de Lacan (voir le chapitre 8 de MRV). Par ailleurs, comme le dénonce Philippe Breton (1992), autour de l'idée que nous sommes des machines et autour de l'utopie de la communication, se sont développées des conceptions passablement nihilistes, comme celle d'un homme « sans intérieur ». Je ne dis pas ici qu'il faille considérer que l'homme a un « intérieur », mais proclamer positivement qu'il n'en a pas serait une plus dangereuse erreur, remettant en cause des siècles d'humanisme. En ce sens, la position du Moi virtuel est effectivement compatible avec les traditions

(hindouiste et bouddhiste) du *neti neti* (ni ceci, ni cela), auxquelles se réfère Varela, et elle évite d'exprimer sous forme affirmative les thèses « postmodernes » ou « déconstructionnistes ». Elle constitue donc aussi une échappatoire au nihilisme.

Certains diront encore : cette conception du virtuel est bien étroite. Elle ne permet pas de parler d'entreprise virtuelle, de société virtuelle ou de relations sociales virtuelles. Réponse : tant mieux, car il n'y a justement pas d'entreprise virtuelle, pas de société virtuelle, pas de relations sociales virtuelles. C'est un mythe dangereux, et ce vocabulaire ne peut que contribuer à l'entretenir. Une entreprise de télétravail est une entreprise bien réelle en tant qu'entreprise. Les relations sociales via Internet ou dans les jeux en réseau sont sans doute appauvries, mais elles restent bien réelles. De même, l'agent artificiel avec lequel nous communiquons peut être virtuel, le monde avec lequel nous interagissons peut être virtuel ; mais cette communication ou cette interaction, ainsi que tout ce qui en résulte dans l'ordre du vécu, sont parfaitement réelles. Cela ne devrait pas surprendre, si l'on songe à l'analogie optique suivante : les rayons réfléchis par un miroir, bien qu'« issus » du reflet virtuel, sont tout à fait réels. Cette comparaison, qui ne fait sens qu'en vertu de notre définition préalable d'une catégorie générale du virtuel, montre au passage qu'introduire celle-ci n'aura pas été un pur exercice de style sans conséquences.

Pour terminer, montrons cependant comment notre définition générale du virtuel peut être appliquée pour justifier certains effets *rhétoriques* non analysés ici. Pour déterminer le sens du mot virtuel dans un usage rhétorique, nous proposons la démarche suivante : rechercher quelles sont les propriétés qui, dans le substantif qualifié de virtuel, permettraient de légitimer cet usage. Autrement dit, rechercher quel aspect du substantif le qualificatif vise à mettre en valeur en niant la réalité des autres aspects.

Exemple (on trouvera d'autres exemples dans MRV) : si je dis que je bois un café virtuel avec mes collègues le midi, ce café n'est certes pas virtuel au sens strict que nous avons défini ici. Notons qu'il n'est pas non plus imaginaire : ni moi ni aucun de mes collègues (à ma connaissance) ne l'a jamais visualisé devant moi, et j'ai encore moins fait semblant de soulever la tasse. Il n'est pas non plus potentiel, car, chacun sachant que je n'en bois pas, il n'a aucune chance d'apparaître devant moi. Quel est alors le message que je veux faire passer ? Simplement ceci : je ne considère pas fondamentalement le café comme une boisson mais comme un acte social, en ce sens que la présence effective d'un café n'est pas une condition essentielle de la pause-café. De ce non-café, que reste-t-il alors qui pourrait se raccorder à notre définition générale ? Il reste que je suis bien assis devant une table, face à un vide qui pourrait être occupé par un café réel, vide qui interrompt la séquence des tasses réelles et qui définit ma place autour de la table. Comme une image virtuelle, ce vide organise notre répartition autour de la table.

## **6. Conclusion**

Nous voulions dégager des concepts et des idées « universels » dans le monde des TIC, qui donneraient aux sciences sociales une meilleure prise sur les « objets techniques » qu'il produit. *Présentation de l'intelligence artificielle « symbolique » comme visant à mettre en place une interopérabilité sémiotico-cognitive entre l'homme et l'ordinateur (qui prolonge l'interopérabilité technique entre machines et logiciels) ; liens conceptuels étroits de cette IA symbolique avec le structuralisme, tant au niveau des systèmes de signes que des principes fondamentaux de constitution du savoir ; dépendance étroite entre ces deux premières idées ; présentation de la réalité virtuelle comme visant à mettre en place une interopérabilité sensori-motrice entre l'homme et*

*l'ordinateur ; définition précise, et compatible avec ses avatars techniques, de la catégorie (ontologique) générale du virtuel ; identification des agents artificiels de l'IA comme agents virtuels ; relation étroite entre les concepts de virtuel et d'interopérabilité : tels sont les points principaux sur lesquels reposent notre manière particulière de penser notre relation à la machine.*

En restant au plus près des nouvelles technologies de l'information et de la communication, nous avons montré que les ressorts par lesquels elles peuvent avoir un impact sur la société et la culture sont beaucoup plus subtils qu'on ne pouvait l'imaginer par les approches sociologiques « externes » habituelles en termes d'usages, d'impact social ou de relations à l'imaginaire collectif ; par exemple, la puissance du virtuel a des sources plus profondes que cet imaginaire. Cette approche « au ras des techniques » conduit en outre à dissoudre certains problèmes et certains discours basés sur des concepts habituellement mal définis, comme celui de virtuel, source de beaucoup de confusion.

Bien sûr, la démonstration de l'utilité effective de nos concepts pour les sciences sociales reste en grande partie à faire. D'ores et déjà, cependant, il est facile de voir que notre définition du virtuel comme pleinement actuel explique sa capacité à se saisir de nous ; nous avons vu aussi qu'elle interpelle directement notre conception du Moi. Quant au concept d'interopérabilité, beaucoup plus que celui généralement utilisé d'interface, il soulève la question de la place que nous voulons faire aux agents et aux mondes virtuels ainsi qu'à l'augmentation du monde réel ; il ne fait pas de doute qu'il pourrait être étendu à d'autres contextes : on pourrait, par exemple, évoquer les organes artificiels, qui soulèvent certes une question d'« interfaçage » physique (travail du chirurgien), mais qui soulèvent surtout des questions de biocompatibilité et de co-fonctionnement ou co-opération avec le reste du corps, questions que le concept d'interopérabilité nous semble mieux subsumer que celui d'interface.

Notons pour terminer une conséquence « politique » évidente de ces considérations, qui mériterait sans doute d'être étudiée pour

elle-même. La possibilité d'interopérabilité avec les agents virtuels de l'IA ou dans les mondes virtuels de la RV a une précondition évidente, qui est un certain « clonage » sur l'ordinateur de nos savoirs, de nos langages et de nos principes de modélisation du monde – clonage « constructif » dans le sens où il est bien connu que développer une base de connaissances a un effet en retour sur le domaine de connaissance concerné. Comme en génétique, ce processus de clonage impose qu'on en envisage toute la dimension culturelle, sociétale, éthique. On concevra aisément que celui qui clone ainsi sa culture (ou une partie) est le mieux à même de la diffuser. Or, à l'autre bout, du côté des utilisateurs, fréquenter des agents virtuels qui dialoguent et raisonnent d'une certaine manière peut avoir des effets en retour très subtils, atteignant aux principes mêmes de constitution de nos savoirs et conduisant à un certain formatage de la pensée ; de même, fréquenter des mondes virtuels conçus d'une certaine manière peut avoir des répercussions sur notre manière de concevoir le monde environnant. On pourra alors se lamenter sur l'hégémonisme américain et l'incapacité française ou européenne à mettre sur pied des programmes de même envergure dans ces domaines.

### ***Bibliographie***

- Berthier Denis, L'agent rationnel abstrait, objet de l'IA ?, *Revue d'intelligence artificielle*, Vol 8, n° 4, 1994, pp 327-359.
- Berthier Denis, *Le savoir et l'ordinateur*, L'Harmattan, coll. « Impacts des Nouvelles Technologies », Paris, 2002.
- Berthier Denis, La culture et la rationalité modernes, du structuralisme à l'IA symbolique, *Automates Intelligents*, n° 42, Paris, Juin 2003.
- Berthier Denis, *Méditations sur le réel et le virtuel*, L'Harmattan, Paris, 2004.
- Breton Philippe, *L'utopie de la communication*, La Découverte, Paris, 1992 ; 2<sup>ème</sup> éd. 1997.
- Breton Philippe, *A l'image de l'homme*, Editions du Seuil, Paris, 1996.

- Deleuze Gilles, *Différence et répétition*, Presses Universitaires de France, Paris, 1968.
- Dupuy Jean-Pierre, *Aux origines des sciences cognitives*, La Découverte, Paris, 1994 ; repris dans la coll. « Poche », Paris, 1999.
- Lacan Jacques, *Le Séminaire, livre I, Les écrits techniques de Freud*, Editions du Seuil, Paris, 1975 (Séminaire de l'année 1953-1954).
- Lafontaine Céline, *L'empire cybernétique. Des machines à penser à la pensée machine*, Editions du Seuil, Paris 2004.
- Mugur-Schächter Mioara, *Quantum Mechanics. Cognition and Action. Proposals in Epistemology*, Kluwer Academic Press, 2002.
- Newell Allen, Physical Symbol Systems, *Cognitive Science*, Vol 4, pp. 135-183, 1980.
- Newell Allen, The Knowledge Level, *Artificial Intelligence*, Vol 59, pp 87-127, 1982.
- Varela Francisco, *Quel savoir pour l'éthique ?*, La Découverte, Paris, 2004.