

« LA SCIENCE COGNITIVE : LA PLUS NOUVELLE DES SCIENCES DE L'ARTIFICIEL »

par Herbert A. SIMON (1980)

Traduction de "[Cognitive Science:: The Newest Science of the Artificial](#)" publié en 1980 dans le Journal of the COGNITIVE SCIENCE, vol 4, (1980) N°1, p 33-46 3346 (1980), traduction assurée par Jean Pierre PAILLET que l'équipe d'animation du Réseau Intelligence de la Complexité MCX-APC remercie chaleureusement pour son précieux concours.

Le champ de cognition n'est-il pas immense, enchevêtrant de nombreux processus en reliances et en boucles, de l'attention-perception-symbolisation-interprétation à la mémorisation-communication-conception-décision ? Science de la cognition assumant sa plénitude se formant dans les sciences de la vie autant que dans les sciences d'ingénierie systémique qui, à la fois empiriquement et pragmatiquement, la forment ?

Défi épistémologique, celui de '*La connaissance de la connaissance*', Défi synthétisant toutes celles que doivent relever les 'nouvelles sciences' se formant désormais dans l'aspiration civilisatrice à l'inter ou à la trans disciplinarité. Défi qu'elles relèvent en redéployant les sciences d'ingénierie (hier dites *sciences appliquées*, souvent encore tenues pour auxiliaires) s'exerçant à leur critique épistémologique interne de façon réfléchie et argumentées et par là pouvant être tenue pour aussi fondamentales que les sciences de la nature

N'est-il pas significatif que cet article, parfois encore tenu pour provocant, ait été publié en 1980 dans le numéro du [journal of the Cognitive Science Society \(Vol 4 n°1\)](#) qui annonçait la constitution formelle de la *Cognitive Science Society* ? Puis que neuf ans après, HA Simon ait publié un gros article intitulé '[Foundations of Cognitive Science](#)' 1989) introduisant un volume collectif (888 pages) publié [sous le même titre](#) (avec CA Kaplan, p 2-47)

Il faut enfin préciser deux options de traduction. L'une est passive : traduire selon l'usage en français l'anglo saxon '*cognitive science*' par '*science cognitive*' alors qu'il serait plus significatif de le traduire par '*science de la cognition*' : Serait-il judicieux de dire de '*la science de la décision*' qu'elle est '*la science décisive*' ? ou de '*la science de l'information*' qu'elle est '*la science informative*' ? On s'est ici résigné à se conformer à l'usage encore très dominant en français (en 2018)

L'autre est active dans ce contexte : traduire le substantif '*The Learning*' par '*L'Apprenance*' plutôt que par '*L'Apprentissage*'. Ce dernier mot, bien que plus usuel dans les traductions de '*Learning*', est 'chargé sémantiquement de l'image du rapport de l'apprenti avec son maître ou tuteur (recopier ou appliquer fidèlement, tacitement, sans chercher d'abord à comprendre). Comme depuis les années 1980 le terme '*Apprenance organisationnelle*' s'est formé pour traduire '*Organisational Learning*' ,chargé d'une image d'interaction récursive, attentive au processus (apprenant) plutôt qu'au résultat (appris), on a pris le parti de le retenir.

Ceci délibérément, à l'heure où en France au moins, *l'essor de l'apprentissage automatique* semble fasciner quasi exclusivement les académies autant que les marchés industriels s'attachant à [définir le sens qu'elles veulent donner à l'Intelligence Artificielle](#) et par là à la science de la cognition, en privilégiant 'l'Apprentissage Automatique', ce subreptice glissement sémantique ne sera peut être pas inutile ? (Note de JL LM 19 06 18).

## « LA SCIENCE COGNITIVE : LA PLUS NOUVELLE DES SCIENCES DE L'ARTIFICIEL »

7 par **HERBERT A. SIMON** 1980 (Carnegie-Mellon University)

Bien sûr, la science cognitive n'est pas vraiment une nouvelle discipline. Il s'agit tout simplement de reconnaître qu'il existe un ensemble de questions qui intéressent en commun de nombreuses disciplines : psychologie, informatique, linguistique, économie, épistémologie, et en fin de compte toutes les sciences sociales. Ces disciplines s'occupent toutes de systèmes qui traitent l'information, et de systèmes adaptifs – c'est-à-dire de systèmes dont la nature résulte des contraintes combinées de leur substrat physique (matériel ou physiologie) *par-dessous*, et de leur environnement complexe *par-dessus*. Un système adaptif peut aussi bien être qualifié d'*artificiel*, dans la mesure où il peut évoluer lorsque son environnement change, comme s'il avait été conçu de propos délibéré pour sa place dans cet environnement (et c'est parfois le cas).

La science empirique a pour tâche de découvrir et tester les invariants des phénomènes qu'elle étudie. La nature artificielle des systèmes de traitement de l'information introduit un problème subtil quant à définir leurs invariants empiriques. Les régularités observées ne représentent un invariant que dans une plage limitée de variation de l'environnement : un énoncé correct de loi pour un tel système doit prendre en compte leur dépendance envers certains traits de l'environnement. On se trouve souvent, par exemple, en psychologie expérimentale, étudier un phénomène d'ordre sociologique (l'influence du passé des sujets) alors que l'on pensait étudier un fait physiologique (la manifestation des propriétés du système nerveux humain). De même, ce n'est que récemment que les économistes étudiant le cycle commercial ont pris conscience de la mesure dans laquelle les paramètres du système dépendent de l'expérience des événements économiques qu'ont vécu les membres de la génération précédente. Pour les sciences de l'artificiel, la séparation entre le normatif et le descriptif n'est pas fortement marquée. Par exemple, en économie, le « principe de rationalité » apparaît parfois comme un invariant descriptif affirmé, parfois comme un conseil normatif offert aux décideurs. De même, en psychologie, les processus adaptifs (l'apprenance) ont toujours occupé une place centrale, et à une certaine époque dominaient toute la recherche. La linguistique aussi a souffert de la confusion du normatif et du descriptif dans l'approche de son sujet. Il nous faut, dans l'étude des systèmes de traitement de l'information, nous garder de croire que les processus adaptifs sont invariants : nous aurons à faire face à une récréation complexe, où les processus adaptifs sont eux-mêmes sujets au progrès adaptif.

Dix ans plus tôt, il y aurait eu lieu d'argumenter sur les points communs des processus de traitement de l'information mis en œuvre dans des systèmes aussi distincts que les ordinateurs et le système nerveux humain. Ceux-ci sont maintenant abondamment démontrés. Ce qui reste en question quant aux limites de la science cognitive est plutôt de découvrir s'il y a des points communs non triviaux entre ceux-ci et les systèmes génétiques.

Nous sommes ici assemblés pour le baptême d'un domaine de recherche scientifique que l'on peut appeler Science Cognitive. Il est bien rare qu'un baptême se fasse attendre si longtemps : selon un estimé de date de naissance des plus conservateurs, la science cognitive a bien longtemps passé l'âge de voter. La preuve convaincante existe que leur naissance remonte au moins à 1956.

Cette cérémonie de baptême ne va pas constituer une nouvelle discipline, mais plutôt offrir les moyens de reconnaître l'existence de préoccupations communes à de nombreuses disciplines : psychologie de la cognition, Intelligence artificielle, linguistique, philosophie, et autres études cherchant à comprendre l'esprit humain. C'est là sans nul doute un objectif louable de réflexion et de recherche. L'histoire de cet effort remonte au moins à Aristote.

S'il en est ainsi, pourquoi mentionner 1956 ? L'importance de cette année émerge du fait qu'elle a signalé une nouvelle manière de comprendre l'esprit humain, un paradigme scientifique nouveau, que nous appelons maintenant le paradigme du traitement de l'information. En 1956 George Miller publiait une explication, en termes de traitement de l'information, des limites de capacité de la mémoire à court terme (Miller 1956) ; Chomsky publiait une de ses premières analyses des propriétés formelles des grammaires transformationnelles (Chomsky 1956) ; Bruner, Goodnow et Austin présentaient les stratégies comme structures de médiation en théorie de la cognition, dans leur ouvrage *A Study of Thinking* ; Allen Newell et moi-même présentions une description du Logic Theorist, premier programme capable de résoudre certains problèmes par recherche heuristique, à l'imitation des humains (Newell & Simon 1956). Quelle année, ce 1956 !

Prenons donc 1956 comme année de naissance de la Science cognitive – de l'analyse de l'esprit humain en termes de traitement de l'information. Dans les années suivantes, elles ont connu une croissance régulière et assez rapide, qu'elle soit mesurée par l'effort de recherche dans le domaine, par les nouvelles connaissances produites sur l'esprit, ou par le degré d'acceptation de la nouvelle approche dans les disciplines concernées.

La croissance de la Science cognitive a engendré de nouvelles revues scientifiques dans plusieurs disciplines, lorsque les publications existantes étaient trop remplies, ou trop conservatrices, pour accepter les nouvelles contributions. Pour n'en citer que deux : *Artificial Intelligence* et *Cognitive Psychology*. Toutefois ces revues restaient confinées à leurs disciplines distinctes ; ce n'est qu'avec le lancement de *Cognitive Science*, il y a trois ans, qu'est apparu un nouveau canal de communication qui délibérément ignore les frontières interdisciplinaires.

## L'ÉTUDE DES SYSTÈMES INTELLIGENTS

Mais ma définition de la Science cognitive est déjà trop étroite. Je donnais l'étude de l'esprit humain comme son objet de recherche. Ceci n'est pas inexact du point de vue historique. Jusque récemment on associait le concept d'intelligence avec le cerveau et l'esprit, particulièrement l'esprit humain. Mais nous apprenons, par la recherche en intelligence artificielle et la simulation de la pensée humaine sur ordinateur, comment construire des systèmes intelligents qui ne sont pas humains, et comment faire abstraction des substrats, cerveaux ou enceintes électroniques, pour identifier les exigences et les marqueurs caractéristiques de l'intelligence.

Je crois donc que la plupart d'entre nous, aujourd'hui, préféreraient définir la Science cognitive comme le domaine de recherche qui vise à comprendre les systèmes intelligents et la nature de l'intelligence. Nous savons maintenant que l'intelligence ne réside pas dans une substance – protoplasme, verre, câblage – mais dans les formes incarnées dans ces substances et dans les processus qui s'y déroulent. A la racine de l'intelligence il y a les symboles, leur capacité de dénotation et leur aptitude à être manipulés. On peut fabriquer des symboles à partir de tout ce qu'on peut organiser et combiner en schémas. L'intelligence consiste en la construction d'un esprit à partir de n'importe quel genre de matière organisable.

Je ne vais pas m'attarder sur le sujet du fondement de l'intelligence sur les structures de symboles et leur manipulation : c'est un thème qui sera élaboré par mon collègue Allen Newell plus tard au cours de ce congrès<sup>1</sup>. Mais dans la suite de ma communication je vais simplement prendre comme acquis ce fondement des comportements intelligents, m'adressant plutôt à un autre aspect des systèmes intelligents que nous connaissons, savoir leur malléabilité et leur adaptabilité, et par conséquent leur caractère fondamentalement artificiel.

### MALLÉABILITÉ ET ADAPTATION

L'intelligence d'un système intelligent se manifeste par le fait qu'ils réalisent leurs objectifs (p.ex. satisfaire leurs besoins de survie) dans des environnements différents et variables. Les comportements intelligents sont adaptatifs, et par conséquent prennent des formes remarquablement différentes dans des environnements différents. La nature d'un système intelligent émerge des contraintes qu'il subit de la part de son substrat physique (qu'il soit matériel ou physiologique) qui limite (par-dessous) son adaptabilité, et de son environnement complexe qui (par-dessus) impose des changements.

Un système adaptatif peut aussi bien être qualifié d'artificiel, dans la mesure où il peut évoluer lorsque son environnement change, comme s'il avait été conçu de propos délibéré pour sa place dans cet environnement (et c'est parfois le cas) (Simon 1969). La science empirique a pour tâche de découvrir et tester les invariants des phénomènes qu'elle étudie. La nature artificielle des systèmes de traitement de l'information introduit un problème subtil quant à définir leurs invariants empiriques. Les régularités observées ne représentent un invariant que dans une plage limitée de variation de l'environnement : un énoncé correct de loi pour un tel système doit prendre en compte leur dépendance envers certains traits de l'environnement.

On se trouve souvent, par exemple, en psychologie expérimentale, étudier un phénomène d'ordre sociologique (l'influence du passé des sujets) alors que l'on pensait étudier un fait physiologique (la manifestation des propriétés du système nerveux humain). De même, ce n'est que récemment que les économistes étudiant le cycle commercial ont pris conscience de la mesure dans laquelle les paramètres du système dépendent de l'expérience des événements économiques qu'ont vécu les membres de la génération précédente (Simon, 1979a).

La découverte d'invariants dans un système artificiel est donc une tâche ardue. Mais il n'y a pas lieu pour autant de désespérer de la Science Cognitive. L'invariance absolue est très rare dans la nature, excepté peut-être la structure de l'espace-temps et les hypothétiques particules élémentaires que les physiciens recherchent si ardemment. La biologie étudie des systèmes qui ne sont apparus

que très tard dans l'histoire de notre système solaire ; si la vie existe ailleurs dans l'univers – comme on nous en affirme la nécessité sur raisonnement probabiliste, il n'est absolument pas assuré que les régularités observées sur terre seront applicables à des formes de vie dans d'autres galaxies.

Ce que nous recherchons, ce sont des invariants relativisés : des régularités applicables sur de grandes étendues temporelles à une gamme importante de systèmes. La définition d'un invariant pour un système adaptif va dépendre de l'étendue temporelle de notre observation. Nous pouvons considérer au moins trois échelles temporelles pertinentes pour ces systèmes, associées à trois formes différentes d'adaptation.

A l'échelle la plus brève, un système intelligent – donc adaptif – modifie constamment son comportement tandis qu'il cherche à résoudre une situation problématique. Par conséquent, une caractéristique majeure de toute recherche heuristique – d'autant plus que cette recherche réussit – est que le système adopte progressivement la forme et le comportement nécessaires pour s'adapter à son environnement.

A une échelle plus longue, un système intelligent produit des adaptations qu'il conserve et qui restent disponibles pour faire face à de nouvelles situations : il apprend. Ces adaptations semi-permanentes peuvent prendre de nombreuses formes, associées à diverses formes d'apprenance. Une de ces formes est l'accumulation d'information en mémoire, et l'acquisition de chemins d'accès pour retrouver cette information. L'apprenance introduit une source de variation énorme dans le comportement d'un système, ce qui complique la recherche d'invariants.

A l'échelle la plus vaste, les systèmes intelligents évoluent. Il peut s'agir d'un processus darwinien, par mutation et sélection naturelle dans le cas des organismes biologiques. Le processus peut aussi être social, par la découverte de connaissances ou de stratégies nouvelles, et leur transmission d'un système à un autre. Ce patrimoine transmis biologiquement ou socialement va aussi causer un changement progressif du comportement, et par conséquent rétrécir le domaine d'invariance.

Au vu de ces capacités d'adaptation, d'apprenance et d'évolution, quel espace nous reste-t-il pour une science générale de la cognition ? Quels sont ces invariants que nous recherchons ? Il nous faut les chercher dans les environnements, interne et externe, qui forment les conditions aux limites du processus d'adaptation. Nous devons nous demander s'il existe des caractéristiques fondamentales communes aux diverses formes de « matériel » intelligent ; nous devons aussi demander s'il existe des caractéristiques communes aux environnements complexes qui font problème.

## L'ENVIRONNEMENT INTERNE

Ce que nous savons des invariants de l'environnement interne des systèmes intelligents se limite à ceux dont nous avons l'expérience. D'une certaine manière ceci se réduit aux organismes vivants et aux ordinateurs. Tous les organismes vivants font usage du même matériau protoplasmique, mais dans des organisations très variées. Tous les ordinateurs construits au cours de l'histoire bien courte de cette espèce ont des organisations remarquablement semblables, mais font usage d'une grande diversité de composants. Autrement dit, dans ces deux classes de systèmes intelligents, nous trouvons une diversité considérable tant de substrat matériel que d'organisation.

Les propos qui suivent sont concentrés sur l'intelligence humaine et l'intelligence informatique. Dans certains cas, j'en viendrai à prendre en compte les insectes sociaux et les organisations humaines. Je ne pense pas que la perspective changerait beaucoup si nous considérons d'autres formes d'intelligence.

Ces deux classes de systèmes sont des systèmes de symboles. Leur intelligence émerge de la symbolisation d'événements et de situations tant internes qu'externes, et de la manipulation de ces symboles. Tous ces systèmes mettent en œuvre les mêmes processus de manipulation symbolique. Il est possible que cet invariant soit dû au fait que les ordinateurs ont été conçus (sans intention déclarée) à l'image de l'être humain. Mais nul n'a jusqu'ici réussi à inventer un système intelligent qui fasse usage de processus exotiques de manipulation symbolique ; peut-être cet invariant est-il plus profond qu'une simple imitation.

Aucun de ces systèmes ne fait preuve de parallélisme important dans son fonctionnement. Bien sûr, cette affirmation prête à controverse dans le cas de l'intelligence humaine ; je ne propose pas de la défendre en détail (même si je la crois correcte et appuyée par les faits)<sup>ii</sup>. Pour les ordinateurs, il y a apparemment des contre-exemples, comme l'ILLIAC-IV, mais on connaît bien la grande difficulté de programmer ces ordinateurs en parallèle.

On peut faire l'hypothèse que la raison principale de cette prévalence de processus séquentiels provient de la grande difficulté à organiser des systèmes de calcul exigeant une coordination précise des calculs exécutés simultanément par divers composants. Cette difficulté, qui a mis en échec les programmeurs humains, semble avoir aussi mis en échec les processus d'apprenance et d'évolution.

Dans un processus essentiellement séquentiel, il est possible de gérer tous les intrants et extrants assez volatiles dans une mémoire active de capacité limitée. Dans le système humain, cette mémoire active se manifeste dans le phénomène bien connu de l'attention. Pour les systèmes d'ordinateurs, nous n'avons pas une semblable limitation, sinon par choix. Pourtant nous voyons émerger des architectures similaires dues à l'arbitrage entre des composants rapides mais coûteux (donc réduits) et d'autres plus lents mais moins coûteux (donc plus répandus).

Il y a aussi l'arbitrage entre la flexibilité d'adaptation et une cohérence d'attention à l'objectif, qui suggère des mécanismes de contrôle de l'attention. En conséquence, dans la mise en place de l'intelligence pour des systèmes de production, on désigne une partie de la mémoire comme mémoire active, pour laquelle les préconditions des règles doivent être satisfaites.

Nous n'avons pas réussi à concevoir des systèmes intelligents capables de prendre en compte simultanément tout ce qui se passe (dans leur mémoire et dans leur environnement) ; il se peut que la nature n'ait pas non plus réussi sur ce point. Si c'est le cas, voici un invariant qu'il y a lieu de spécifier et de comprendre.

Mais n'est-ce pas ignorer toute l'activité actuelle sur les multiprocesseurs, qui sont la quintessence du parallélisme ? Et ignorer aussi d'autres vénérables systèmes parallèles, que sont les organisations humaines et les fourmilières ? Je crois pouvoir résoudre cette contradiction apparente. Je parlais de la grande difficulté à organiser des systèmes « exigeant une coordination précise des calculs exécutés simultanément par divers composants. » Le secret des organisations humaines et des

fourmilières est qu'elles n'exigent pas de coordination précise entre les employés humains, ou les individus fourmis.

La durée des processus de pensée peut se mesurer en millisecondes, en secondes ou en minutes. La coordination des comportements des membres d'une organisation humaine n'exige pas de transmettre l'information à la milliseconde, ni même en général à la seconde ou à la minute. Sans doute les limites de précision sont mises à l'épreuve dans les équipes sportives, mais même dans ce cas, la quantité d'information à transmettre est infime comparée à la vitesse de traitement dans la tête de chaque individu.

Naturellement, si chaque membre de l'organisation s'adonne à une tâche différente, ou si les besoins de coordination sont modestes, rien n'empêche le parallélisme. Tous les exemples de parallélisme dans la nature semblent conformes à ce principe général : le taux d'interaction entre composants est faible par rapport au taux d'interaction à l'intérieur de chaque composant. Les systèmes qui ont cette propriété sont qualifiés de quasi-décomposables, et cette quasi-décomposabilité entraîne plusieurs conséquences théoriques intéressantes (Simon 1969, chapitre 4 ; Courtois 1977).

Je fais cette prédiction : notre progrès dans la conception de multiprocesseurs de plus en plus grands et complexes va nous faire voir de plus en plus clairement dans leur architecture cette quasi-décomposabilité, avec la quasi-hiérarchie qui l'accompagne<sup>iii</sup>. Ceci va se produire, que les concepteurs s'inspirent des organisations humaines (ou des fourmilières), ou ignorent complètement les documents de cette expérience et choisissent de « réinventer la roue. »

Ces observations suffiront peut-être à donner une idée de la nature des invariants que nous pouvons espérer découvrir dans l'environnement interne des systèmes intelligents. Mes exemples suggèrent que ces invariants sont relativement abstraits, tendant à imposer des contraintes sur l'organisation d'un système plutôt que sur le substrat matériel. Il semble qu'un système intelligent doive être un système symbolique, dont les composants de haute fréquence doivent fonctionner séquentiellement, et être doués d'attention sélective, tandis que les systèmes plus complexes seront hiérarchiques et quasi-décomposables.

## L'ENVIRONNEMENT EXTERNE

Le deuxième type d'invariance dans les systèmes intelligents provient des caractéristiques communes aux environnements dans lesquels ils fonctionnent, et auxquels ils doivent s'adapter. Ici encore, vu la diversité de tels environnements les invariants seront de nature très abstraite.

Les environnements qui nous intéressent, ceux qui forcent l'exercice de l'intelligence, sont des environnements problématiques. Ils n'offrent pas d'itinéraires évidents pour atteindre les objectifs du système. Souvent (mais pas toujours) un environnement problématique contient un grand nombre (parfois une immensité) de possibilités, dont seule une faible fraction pourra satisfaire les exigences de l'objectif.

Le principal mécanisme d'intelligence observé (chez les ordinateurs comme chez les humains) dans un environnement problématique est la recherche heuristique. La partie « recherche »

de ce processus est assez claire. Il y a plus de subtilité dans les moyens heuristiques qui permettent une meilleure sélectivité dans la recherche :

- (1) Faire usage d'information en mémoire pour choisir des itinéraires plus prometteurs
- (2) Tirer de l'environnement du problème des informations sur les régularités de structure qui pourront aussi guider la recherche.

Comme nous avons acquis dans ce dernier quart de siècle beaucoup de connaissances et une documentation considérable sur la recherche heuristique, je ne vais pas m'attarder sur ce sujet (voir Nilsson 1971 pour référence).

Avant de quitter ce sujet, je veux toutefois attirer l'attention sur l'importance critique de l'interface entre l'environnement et le système intelligent, savoir les organes sensoriels et moteurs de ce dernier. Cette interface représente ce qui, pour multiples raisons, constitue le problème le plus délicat dans la conception d'un système intelligent : elle doit satisfaire simultanément aux contraintes des environnements, interne et externe – en fait soutenir la communication entre eux. Il n'est sans doute pas accidentel que nous ayons fait beaucoup moins de progrès dans la conception de capteurs et d'effecteurs, ou dans l'imitation des organes humains, que dans notre compréhension et imitation des processus d'intelligence qui se déroulent à l'intérieur du crâne humain ou de la boîte noire.

### L'APPREANCE

Ayant reconnu que les systèmes intelligents ont des comportements extrêmement adaptables et flexibles, on pourrait en arriver à rechercher leurs invariants, non dans leur comportement ou la structure qui soutient leur performance, mais dans les mécanismes à long terme qui produisent leur adaptation, autrement dit, les mécanismes d'apprenance

De fait, l'apprenance fut un sujet favori au début de la recherche en intelligence artificielle, et en psychologie, il a été le sujet quasi-dominant depuis la première Guerre Mondiale jusqu'au milieu des années Cinquante. Nous ne voyons pas clairement les raisons de cette prédilection, et nous ne saurions prendre pour acquit que la raison mentionnée au dernier paragraphe a été prépondérante.

Quant à l'intelligence artificielle, j'ai l'impression que bien des chercheurs pensaient qu'il serait plus facile d'amener un système à s'auto organiser à partir d'une séquence bien choisie d'expériences d'entraînement que de lui inculquer les connaissances nécessaires à un fonctionnement d'expert. Certaines attitudes étaient peut-être transmises de la psychologie, qui plaçait alors l'apprenance au centre de son sujet, alors que Hebb et d'autres étudiaient la possibilité d'auto organisation des réseaux de neurones<sup>iv</sup>. Il y avait enfin sans doute un désir d'échapper à l'accusation souvent dirigée contre les systèmes d'intelligence artificielle, que l'intelligence résidait en fait dans le programmeur, et non dans le système. Cette accusation n'avait plus cours si le programmeur fournissait seulement le potentiel d'apprenance.

Quelle qu'en ait été la raison, bien des chercheurs ont adopté le chemin de l'apprenance, il est correct, je crois de dire que dans l'ensemble ils ont échoué. Tous les systèmes experts produits jusqu'ici ont été dotés directement des connaissances et des stratégies de solution qui forment leur expertise.



Il faut prendre soin de ne pas extrapoler cette expérience des deux décennies passées aux deux suivantes. Les cinq dernières années ont vu des signes d'une reprise de l'effort sur l'apprenance, mais dans une perspective tout autre que celle des réseaux de neurones et de simples modèles de renforcement. Des systèmes comme METADENDRAL (Buchanan & Mitchell 1977) qui induisent leurs propres théories, ou AM de Lenat, qui apprennent de nouveaux concepts qui à leur tour les assistent à en découvrir d'autres, Sont certainement des systèmes capables d'apprenance. De même UNDERSTAND (Hayes & Simon 1974) qui produit, et ensuite utilise, des représentations de problèmes ; ou BACON (Langley 1979), qui peut induire des lois scientifiques par des procédures récursives ; ou le programme de (Neve 1978) qui apprend des techniques par analyse d'exemples tirés de manuels ; et bien d'autres.

Le trait marquant de cette nouvelle génération de programmes d'apprenance est que pour la plupart ce sont des programmes de solution de problèmes, capable d'entreprendre des recherches heuristiques au moyen d'un outillage raisonnable de méthodes telles que « générer-tester » ou « l'analyse moyens-fins ». Ils ne partent pas d'une table rase comme les systèmes auto-organisant de la génération précédente.

Je reviens maintenant à cette préoccupation pour l'apprenance qui a dominé la psychologie juste avant la révolution informationnelle. J'ai suggéré plus tôt qu'elle a pu être motivée par l'idée que seuls les paramètres de l'apprenance pourraient fournir des invariants dans un système artificiel. Mais l'examen de l'œuvre des plus importants chercheurs n'appuie pas cette hypothèse. Je trouve à peine une allusion dans la première édition du *Theories of Learning* de Hilgard (1948), et rien du tout dans les ouvrages de Hull, Watson ou Thorndike. Les raisons historiques sont apparemment tout autres. Tout d'abord, il y avait une motivation pratique : comprendre l'apprenance était essentiel pour comprendre l'éducation. Cette motivation apparaît clairement dans les travaux de John Dewey, Thorndike et autres. Mais plus fort encore que l'éducation, il est un objectif philosophique liant la recherche sur l'apprenance aux questions primordiales de la psychologie.

Bien sûr, la psychologie émerge de la philosophie. La question philosophique qui mène le plus directement aux questions psychologiques est la question épistémologique : comment avons-nous connaissance du monde qui nous entoure, et comment l'esprit garde-t-il cette connaissance du monde ? Il est naturel de passer de la question *comment connaît-on le monde* à celle *comment acquiert-on cette connaissance* : comment se fait l'apprenance, (ainsi que la question associée *comment se fait la perception*). Il est plausible que cette question épistémologique ait amené cette préoccupation pour l'apprenance à une époque où la psychologie se détournait (parfois agressivement) des applications.

Je vais maintenant laisser de côté ces questions historiques et montrer une quasi-contradiction dans ce retour à l'apprenance que j'ai dit prendre place au cours des cinq dernières années. Si la plus grande partie de ce que nous avons appris sur l'expertise (tant en psychologie qu'en intelligence artificielle) provient de nos études et réalisations pratiques de systèmes d'exécution, pourquoi en revenons-nous maintenant à une approche « table-rase » ?

Nous avons une partie de la réponse : les nouveaux systèmes d'apprenance ne ressemblent pas beaucoup à ceux d'il y a vingt ans. Dès le départ, ils ont une organisation bien plus complexe et élaborée que ceux de jadis fondées sur les réseaux de neurones et la perception. Nous avons donc de

bonnes raisons de croire qu'ils pourront escalader l'échelle des aptitudes, et déjà l'expérience acquise est à l'appui de cet espoir.

Une autre partie de la réponse est que nous avons maintenant un bien meilleur aperçu de l'objectif de l'apprenance, des systèmes d'exécution experts que l'apprenant cherche à développer. Nous savons bien des choses sur l'enregistrement de l'expertise dans des mémoires associatives en listes, et sur la manière d'organiser des connaissances pour en faciliter l'accès.

Une troisième partie de la réponse est que nos systèmes experts sont maintenant pour la plupart des systèmes de production, et que nous comprenons mieux comment faciliter l'auto-amorçage de tels systèmes (systèmes de production adaptifs, Waterman 1975) que nous n'avons jamais su pour des systèmes organisés en hiérarchies de sous-programmes fermés.

Je ne dis pas que nous allons délaïsser notre concentration actuelle, en psychologie et en intelligence artificielle, sur les systèmes d'exécution, pour l'étude de l'apprenance. Je ne suis pas non plus en faveur de ce changement. Mais comme nous observons augmenter le nombre des chercheurs en Science Cognitive, il y a sans doute assez de mains et de têtes pour couvrir les deux fronts.

Quand nous étudions l'exécution, méfions-nous de voir des invariants où ils ne sont pas. Etant donné que les systèmes intelligents sont programmables, nous devons nous attendre à voir des systèmes (même de la même espèce) exécuter la même tâche au moyen de stratégies très différentes. A ma connaissance, il n'y a pas de théorème déclarant l'unicité d'une stratégie bonne ou idéale. Nous devons donc nous attendre à trouver des différences de stratégie non seulement à des niveaux de compétence différents, mais aussi parmi les experts.

Par conséquent, la recherche sur la performance des systèmes adaptifs doit revêtir un aspect taxinomique, ou même sociologique. Il y a beaucoup à apprendre sur la variété des stratégies : ne dédaignons pas, n'évitons pas la tâche laborieuse, ordinaire de description de cette variété. Le substrat descriptif est tout aussi nécessaire en science cognitive qu'en biologie. Dans le domaine de la Science Cognitive, seule la linguistique (et en quelque mesure la psychologie du développement) a une tradition de description détaillée – par opposition à l'expérimentation à la recherche de vérités générales.

En ce qui concerne l'aspect sociologique, les programmes d'exécution résultent en grande part de l'apprenance sociale : nous ne devons pas penser que les programmes du vingtième siècle seront les mêmes que ceux du dix-neuvième, du dixième ou du premier. Cette perspective de changements majeurs dans les programmes cognitifs humains nous ramène à d'importants problèmes philosophiques qui n'ont pas attiré beaucoup d'attention en recherche psychologique. Sous quels aspects l'esprit d'un citoyen de la Grèce antique est-il semblable ou différent de celui d'un humain moderne ? La pensée dans un village d'Afrique est-elle le même processus que dans une ferme mécanisée. Nous avons des ressources expérimentales pour étudier la deuxième question, sinon la première ; il y a déjà eu un peu de recherche. Mais si nous prenons au sérieux la proposition que les programmes cognitifs humains sont déterminés au moins autant par des effets sociaux et historiques que par la neurologie, l'orientation de notre recherche va sûrement changer.

Pour les sciences de l'artificiel, la séparation entre le normatif et le descriptif n'est pas fortement marquée. Par exemple, en économie, le « principe de rationalité » apparaît parfois comme un invariant descriptif affirmé, parfois comme un conseil normatif offert aux décideurs. L'entreprise est supposée agir pour maximiser son profit, et on élabore des théories de l'économie sur la base de cette hypothèse. Par ailleurs, les spécialistes en recherche opérationnelle, avec leurs outils de programmation linéaire, de programmation en nombres entiers et de théorie des files d'attente, proposent d'aider les entreprises à maximiser leurs profits si elles ne l'ont pas fait.

De même, en psychologie, l'opinion que les systèmes intelligents sont capables d'adaptation, d'apprenance et d'évolution ne nous empêche pas de chercher des moyens d'améliorer leur aptitude à apprendre ou résoudre des problèmes. La contradiction entre le positif et le normatif est moins éprouvée en psychologie qu'en économie, précisément parce que nous supposons rarement que l'adaptation, l'apprenance ou l'évolution mènent à maximiser une quelconque mesure (auquel cas ils ne seraient pas susceptibles de progrès).

L'émergence récente de la sociobiologie, qui prétend expliquer, ou même prédire, l'état des choses par des raisonnements d'ordre adaptatif, risque d'introduire de nouvelles confusions dans le rapport du normatif au positif. Je reviendrai sur ce point dans la section suivante.

La linguistique aussi a souffert de la confusion du normatif et du descriptif dans l'approche de son sujet. Sans même considérer l'antique notion des faiseurs de dictionnaires que leur rôle était de prescrire l'usage « correct », je pense à une idée plus moderne selon laquelle le langage se définit en termes de « compétence » et de « grammaticalité ». Cette attention portée sur une compétence idéale résidant dans un genre de firmament platonique (ou cartésien) risque fort d'imposer des contraintes normatives sur l'étude du comportement linguistique réel.

Je ne m'oppose pas à l'intérêt d'une science normative dans ces domaines, mais tout simplement à la confusion des points de vue normatif et positif : de ce qui est et de ce qui devrait être. En particulier, l'Intelligence artificielle a des objectifs normatifs, comme il se doit. Dans son aspect applicatif, elle cherche non seulement à comprendre l'intelligence, mais aussi à l'améliorer. Peut-être devons-nous le reconnaître explicitement en parlant d'ingénierie cognitive et de science cognitive. Mais dans ce cas, j'espère que ces deux entreprises resteront en relation étroite, comme par le quart-de-siècle passé. Il y a moins à redouter d'une confusion entre normatif et positif que des pertes à encourir par une séparation entre la science et ses applications ingénieriales.

## ÉVOLUTION ET OPTIMISATION

Je reviens maintenant à la théorie de l'évolution, spécifiquement à sa manifestation actuelle en sociobiologie (Wilson 1975). Les économistes en particulier se saisissent des arguments évolutionnistes pour étendre leur domaine au-delà des phénomènes strictement économiques pour couvrir tous les comportements individuels et sociaux (Simon 1979a).

Par son slogan « la survie du plus apte » la théorie de l'évolution semble bien sûr prétendre à l'optimisation. Il importe d'examiner la signification de cette expression, et sa signification possible dans le monde que nous connaissons. « La survie du plus apte » concerne l'issue de la concurrence entre les espèces qui cherchent à occuper une certaine niche écologique. Dans la forme la plus simple de cette théorie, il peut exister au moins autant d'espèces survivantes, adaptées, qu'il y a de niches : un

nombre énorme. De plus les niches elles-mêmes ne sont pas déterminées par un environnement inflexible, invariant, mais en grande partie par la constellation des organismes eux-mêmes. Pas de poux sans têtes chevelues à investir, pas d'animaux sans végétaux. En conséquence, on ne voit pas très bien quel serait le problème d'optimisation résolu dans le processus d'évolution. On peut tout juste dire que l'occupation de chaque niche est « optimisée » localement dans le contexte de la configuration d'ensemble des niches.

Il est deux domaines de formalisation, la théorie des jeux et la programmation linéaire, qui nous fournissent des aperçus importants sur la difficulté qu'il y a à vouloir appliquer des notions d'optimisation pour prédire le comportement des systèmes complexes. La théorie des jeux nous enseigne que dans les situations de concurrence il est parfois impossible de définir un optimum sans ambiguïté, sans même chercher à garantir l'existence ou l'unicité d'une solution optimale. La programmation linéaire nous enseigne qu'il faut des conditions extrêmement restrictives (p.ex., linéarité de la fonction critère, convexité de l'espace de faisabilité) pour garantir qu'un maximum local sera bien un maximum global, et qu'en l'absence de ces restrictions, il est très difficile – en fait, généralement impossible – de trouver un algorithme pour découvrir un maximum global au prix d'un effort de calcul raisonnable.

C'est pour ces raisons que la Science cognitive demeurera probablement une science des systèmes 'satisfaisants'<sup>v</sup> – c'est-à-dire de systèmes qui trouvent des solutions acceptables à leurs problèmes -- plutôt qu'une science des systèmes « optimisants » -- c'est-à-dire de systèmes parfaitement adaptés à leur environnement. Mais pour comprendre l'intelligence '*satisfaisante*', il faut comprendre le processus de l'exercice de cette intelligence. Il n'est pas suffisant d'inférer la nature « nécessaire » du comportement intelligent à partir de la description de l'environnement et des conditions d'optimisation. Je ne vais pas pousser plus loin cet argument, car je crois que la plupart des chercheurs dans mon auditoire sont probablement d'accord sur ce point.

## CONCLUSION

Dix ans plus tôt, il y aurait eu lieu d'argumenter sur les points communs des processus de traitement de l'information mis en œuvre dans des systèmes aussi distincts que les ordinateurs et le système nerveux humain. Ceux-ci sont maintenant abondamment démontrés. Ce qui reste en question quant aux limites de la Science cognitive est plutôt de découvrir s'il y a des points communs non triviaux entre ceux-ci et les systèmes génétiques. Où que se place la limite, il existe maintenant une science des systèmes intelligents qui transcende les frontières de toute espèce unique.

L'intelligence a un rapport étroit avec l'adaptabilité – solution de problèmes, apprenance et évolution. La science des systèmes intelligents doit être une science des systèmes adaptatifs, avec tous les défis dans la découverte d'invariants. La Science cognitive est une science de l'artificiel. Elle s'occupe de phénomènes qui pourraient prendre une forme différente – et qui vont changer continuellement dans leur adaptation aux exigences de leur environnement.

Pourvu que nous ne fassions pas la confusion entre l'adaptabilité et la capacité d'atteindre des solutions optimales, la Science cognitive restera une science essentiellement empirique. Un raisonnement fondé sur les conditions d'optimisation pourra jouer une part modeste dans la découverte de ce qui est. La plus grande partie de nos connaissances résulteront d'une observation

laborieuse de la vaste variété de programmes que ces systèmes acquièrent et utilisent, et de théories formalisées – principalement sous forme de programmes d'ordinateur – produites à partir de ces observations.

## RÉFÉRENCES

- Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.
- Buchanan, B. G., & Mitchell, T. M. Model-directed learning of production rules. Presented at the *Workshop on Pattern-directed inference Systems*, Honolulu, Hawaii, May 1977,
- Chomsky, N. Three models of the description of language. *Proceedings of a Symposium on Information Theory. IRE Transactions on Information Theory*, sept. 1956, IT-2(3), 1 13-124.
- Courtois, P. J. *Decomposability: Queueing and computer system applications*. New York: Academic Press, 1977.
- Hayes, J. R., & Simon, H. A. Understanding written problem instructions. In L. W. Gregg (Ed.), *Knowledge and Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1974.
- Hilgard, E., & Bower, G. H. *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall,
- Langley, P. Rediscovering physics with BACON. 3. *Proceedings of the 6th IJCAI—Tokyo, 1979, I*, 505-507.
- Miller, G. A. The magical number seven. *Psychological Review*, 1956, 63, 81.
- Minsky. M. Steps toward artificial intelligence. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill, 1963.
- Neves, D. A computer program that learns algebraic procedures by examining examples and by working test problems in a textbook. *Proceedings of the 2nd CSCSI Conference—Toronto, 1978*, 191-195.
- Newell, A., & Simon, H. A. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, March 1976, 19, 111—126.
- Newell, A., & Simon, H. A. The logic theory machine. *IRE Transactions on Information Theory*, sept. 1956, IT-2(3), 61-79.
- Nilsson, N. J. *Problem solving methods in artificial intelligence*. New York: McGraw-Hill, 1971.
- Rochester, N., Holland, J. H., Haibt, L. H., & Duda, W. L. Test on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, 1956, IT-2(3), 80-93.
- Simon, H. A. Rational decision making in business organizations. *American Economic Review*, Sept. 1979, 69, 493-513.
- Simon, H. A. *Models of thought*. New Haven: Yale University Press, 1979b.
- Simon, H. A. *The sciences of the artificial*. Cambridge, MA: M.I.T. Press, 1969.
- Waterman, D. A. Adaptive production systems. *Proceedings of the 4th IJCAI – Tbilisi, Georgia, USSR, 1975. 1*:296-303.
- Wilson, E. O. *Sociobiology: The new synthesis*. Cambridge, MA: Belknap Press of Harvard University, 1975.

<sup>i</sup>Voir un traitement initial de cette perspective dans Newell & Simon 1976.

<sup>ii</sup>Je dois bien sûr excepter de mon assertion sur le parallélisme le cas des organes sensoriels, qui sont clairement des appareils fonctionnant en parallèle. Mais une fois exécutée l'extraction des caractéristiques, le processus de reconnaissance de formes peut se faire en temps réel dans des réseaux de discrimination en série. Il est pratiquement certain que ce qui paraît être un traitement parallèle au niveau central (p.ex. converser en conduisant une voiture) est un partage de temps dans un système sériel. Voir (Simon 1979b chapitre 2.3).

<sup>iii</sup>Le mot hiérarchie a ici le sens que je lui donne dans (Simon 1969, chapitre 4, désignant non une structure de contrôle en pyramide, mais une structure modulaire (quasi-décomposable) organisée en « couches » selon la fréquence et la granularité temporelle des interactions entre modules, entre leurs sous-modules, etc. Il est possible pour un système à contrôle « démocratique » d'avoir une organisation hiérarchique en ce sens.

<sup>iv</sup>Voir un exemple important de travail sur ce point de vue dans (Rochester, Holland, Haibt & Duda 1956), et une critique de cette approche dans (Minsky 1963 section III).

<sup>v</sup>NdT H. Simon utilise l'expression "satisficing" (issue du latin Satisfecit) un peu différent de mot « satisfying », qui permet de contraster avec l'expression « optimizing »