

A propos de Philip Johnson-Laird, "L'ordinateur et l'esprit"

© [Jean-Pierre Meunier](#) et André Berten (1995)

La traduction en français de l'ouvrage de Ph. Johnson-Laird, *The Computer and the Mind* (1) constitue d'abord une remarquable présentation, simple et claire —et, de ce point de vue, très utile pour les chercheurs en sciences humaines— des grands concepts et des principaux résultats des sciences cognitives. On y trouvera ensuite un exposé succinct de la théorie des modèles mentaux (2), théorie particulièrement intéressante pour les chercheurs en communication. Il s'agit enfin d'un ouvrage résolument engagé dans le grand débat sur la nature de l'esprit et des opérations mentales, et l'auteur prend ouvertement position en faveur d'un cognitivisme pur et dur.

Commençons par présenter le contenu de l'ouvrage et ses principales articulations.

Telles qu'elles se sont constituées aux États-Unis, suite à diverses influences (la cybernétique, la naissance de l'ordinateur...), les sciences cognitives se sont donné pour objectif d'expliquer le fonctionnement de l'esprit dans un cadre théorique où les notions de symbole et de calculabilité jouent un rôle déterminant. La première partie de l'ouvrage est consacrée à la présentation de ces notions fondamentales. "Le rôle des processus mentaux est de traiter un grand nombre de perceptions, d'idées, d'images, de croyances, d'hypothèses, de pensées et de souvenirs. Un des principes de la science cognitive est que tous ces éléments sont des représentations mentales ou des symboles d'un genre ou d'un autre" (p. 33). L'esprit manipule ces symboles et peut établir un lien entre les symboles obtenus et le monde (comme dans la perception). Comme l'esprit, l'ordinateur est un dispositif manipulant des symboles (des numéraux binaires traduits en impulsions électriques), et c'est cette analogie qui fonde l'idée que les processus informatiques peuvent modéliser les symboles mentaux et leur manipulation.

Les symboles se combinent selon des règles qui constituent des grammaires. Celles-ci sont liées à des programmes dans la mesure où le résultat d'un programme — c'est-à-dire une suite d'instructions — peut être cerné par une grammaire. L'auteur expose les différents types de programmes (déterministe, non déterministe), introduit l'important concept de puissance de calcul et soutient l'hypothèse selon laquelle tous les processus calculables le sont par des "machines de Turing". Un "programme d'ordinateur est semblable à une machine de Turing particulière et l'ordinateur est analogue à une machine universelle parce qu'il peut exécuter n'importe quel programme écrit dans un code approprié." (p. 57)

Dans les pages consacrées à la vision (deuxième partie, chapitres 4, 5 et 6), l'auteur, reprenant, entre autres, les théories de Marr, examine les calculs effectués par le système visuel au cours des différentes phases de la perception. Mais il montre surtout qu'il est impossible de comprendre les problèmes liés à la vision sans une théorie plus large de la connaissance.

Ce problème de la connaissance fait l'objet de la troisième partie. Il conduit naturellement à celui de l'apprentissage, de la mémoire et l'action. Vu comme un exercice de calcul, l'apprentissage consiste à construire de nouveaux programmes à partir d'éléments d'expérience, ce qui implique que l'apprentissage doit dépendre de programmes innés de production de programmes. Après avoir réfuté la thèse sceptique selon laquelle il est impossible d'apprendre des concepts, Johnson-Laird montre aussi que la croyance

op-posée en l'existence d'une procédure universelle capable d'apprendre n'importe quelle tâche est illusoire. L'apprentissage n'est possible que si des contraintes (telles la "grammaire universelle" de Chomsky) limitent la catégorie des concepts ou des compétences à acquérir.

Ces questions de connaissance ou d'apprentis-sage impliquent, comme on le voit, une conception générale de l'esprit Il faut prendre en compte l'architecture globale de l'esprit. Or celle-ci a fait l'objet de deux grands types de théorie: d'après la première, l'esprit serait fondé sur "un système de productions" et d'après la seconde, sur "un traitement parallèle distribué". Ph. Johnson-Laird consacre un chapitre à chacune de ces deux grandes théories et permet ainsi au lecteur de se faire une idée claire de notions fréquemment rencontrées dans la littérature cognitive.

Dans le cadre de la théorie du système de productions, la planification est un élément essentiel de la cognition et consiste en un rappel des connaissances des opérations pertinentes pour atteindre certains objectifs. Ces connaissances seraient représentées dans la mémoire à long terme par un ensemble de règles conditionnelles de la forme:

condition => action

Ces règles font partie d'un programme.

Les systèmes de productions ont servi à bâtir un certain nombre de théories psychologiques, notamment au sujet de la mémoire. Pour les théoriciens des systèmes de productions, les processus qui régissent l'apprentissage et le fonctionnement de la mémoire reposent sur des règles symboliques, d'où certaines difficultés:

L'importance d'un symbole particulier dans un système de productions dépend entièrement des règles dans lesquelles il apparaît, et ces règles doivent être clairement explicitées à l'intérieur du système. Vous pensez peut-être qu'il est plutôt curieux que ces règles n'émergent pas à la surface de votre vie mentale. Si vous pouviez soulever le couvercle de l'esprit et y jeter un coup d'œil, verriez-vous vraiment un ensemble de principes organisés comme les clauses d'une loi, et ces principes cerneraient-ils toutes les éventualités? Peut-être pas. Ces doutes sont en partie à l'origine de l'autre conception de l'architecture mentale (...)
(p. 182).

Cette "autre conception de l'architecture mentale", appelée "traitement parallèle distribué" ou "connexionisme" n'utilise pas de règles ayant une structure explicite et ne conçoit pas la mémoire comme un ensemble de boîtes ayant chacune un contenu particulier. Nous connaissons une capacité remarquable de la mémoire humaine: une chose en évoque une autre; par exemple, les mots ‐ ou les lettres d'un mot ‐ s'appellent l'un l'autre. Ce fait d'expérience suggère un système dont la puissance tient au mode de connexion des unités.

Dans de tels systèmes d'interconnexions, les représentations peuvent être réparties sur de nombreuses unités de mémorisation dont chacune intervient dans la représentation de nombreux éléments. L'apprentissage consiste alors essentiellement à établir, par expérience, les forces de connexion appropriées.

L'intérêt du connexionisme, dit Johnson-Laird, est de montrer que les symboles ne doivent pas forcément être représentés sous forme d'entités distinctes:

Les processus conscients reposent peut-être sur des symboles de ce type, mais les processus inconscients ne peuvent pas être des opérations cachées sur le même genre de symboles. En revanche, ils pourraient bien être le traitement parallèle de représentations distribuées résultant de la fusion de nombreuses expériences différentes (p. 205).

On conçoit l'intérêt de ces propos pour l'approche du vieux problème des niveaux de conscience. On trouvera la même complexité théorique dans le reste de l'ouvrage, et

notamment dans la quatrième partie, consacrée à la cogitation et dans laquelle se trouve présentée l'importante théorie des modèles mentaux mise au point par Johnson-Laird lui-même.

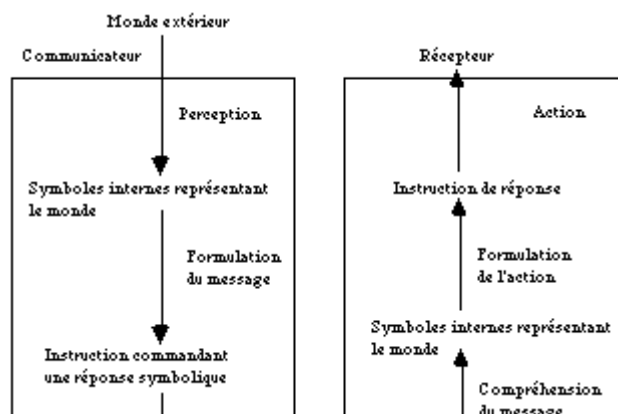
La partie sur la cogitation commence par ce constat: il existe plusieurs types de pensée dont la rêverie —rapide, involontaire, extérieure à la conscience— et le calcul mental —volontaire, conscient, déterministe— seraient deux extrêmes. Le raisonnement suit un objectif précis mais il n'existe pas toujours de procédure toute faite pour l'atteindre.

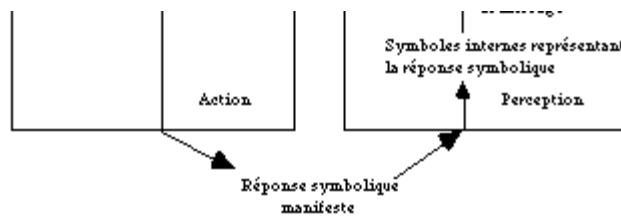
Concernant la déduction dans le raisonnement quotidien, les chercheurs en science cognitive se partagent en deux camps. Pour certains, si les gens sont capables de déductions valides, c'est qu'ils ont des règles formelles dans la tête; cette logique mentale serait une version du calcul des prédicats. Il est cependant plusieurs obstacles à cette théorie formelle du raisonnement humain. En particulier, il semble que le raisonnement humain tienne compte du contenu sémantique des problèmes qu'il tente de résoudre, ce qui est contraire à la notion de règles d'inférences formelles. Les effets de contenu peuvent s'expliquer par l'existence dans l'esprit de règles d'inférence spécialisées (règles d'inférence ayant un contenu spécifique emprunté aux systèmes d'experts humains et qui pourront être représentées dans le mode d'un système de productions). Mais cette hypothèse ne permet plus d'envisager un mécanisme général d'inférence. Il faut donc un compromis entre une capacité générale et une sensibilité au contenu, compromis auquel satisfait la théorie des modèles mentaux (cf. infra).

Concernant l'induction, l'auteur, après avoir examiné les formes possibles de généralisation, examine la nature des concepts ordinaires et rappelle quelques unes des observations les plus intéressantes faites à leur sujet. Les concepts ordinaires ne répondent pas au processus d'abs-traction décrit par les philosophes empiristes. Comme le pensait Wittgenstein et comme tendent à le montrer diverses théories psychologiques contemporaines, les concepts reposent sur des réseaux de similitudes plutôt que sur des éléments communs et ils sont reliés les uns aux autres par des relations hiérarchiques ou d'autres plus complexes.

Faisant suite au chapitre sur l'induction, le chapitre sur la création s'appuie sur l'improvisa-tion musicale ou d'autres genres d'improvisation pour développer une modélisation de la création comme processus à plusieurs phases dont les prémisses reposent généralement sur des contraintes initiales.

La cinquième partie consacrée à la communication constitue une approche ordonnée, dans le cadre d'un modèle général du processus de communication, des différents niveaux que l'on peut discerner: la parole et l'audition, la grammaire et le sens. Le modèle de la communication auquel se réfère Johnson-Laird est proche du modèle désormais classique de Shannon et Weaver en ce sens qu'il a la forme d'une chaîne linéaire. Sous sa forme complète, cette chaîne de la communication symbolique est représentée par le schéma suivant (p. 293):





Le langage humain, le plus puissant des systèmes de communication, comporte une structure hiérarchique dont les différents niveaux (phonèmes, mots, grammaire, sens) obéissent à des règles, même si l'organisation de l'esprit lui permet d'acquérir et de suivre ces règles sans savoir ce qu'elles sont. Procédant par degrés successifs, le chapitre consacré à la parole et l'audition envisage les sons (leur production et leur audition), la linguistique des sons (phonologie), les mots et les problèmes que posent leur identification par un auditeur, l'intonation et le ton de la voix.

La conclusion du chapitre est que de nombreux aspects de la parole sont régis par des principes inconscients dont la nature reste à élucider.

La plupart des analyses théoriques ont supposé que ces principes inconscients sont des règles qui ont une structure explicite. Il se pourrait bien qu'en réalité, ils soient représentés sous une forme distribuée. Le problème reste à résoudre empiriquement (p. 322).

Le chapitre sur la grammaire pose le problème du genre de grammaire qui est représenté dans l'esprit. Johnson-Laird rappelle les théories élaborées par Chomsky, mais il fait remarquer que les règles établies par les linguistes ne correspondent pas aux processus psychologiques ou règles linguistiques tacites mises en œuvre dans la compréhension des phrases. Des études de la mémoire ont montré, par exemple, qu'on se souvient du sens mais pas de la structure syntaxique profonde.

Pour l'auteur les questions principales relatives à la structure syntaxique ont trait à la nature des opérations générales de l'analyseur syntaxique mental, au traitement des ambiguïtés et enfin aux rapports entre l'analyse syntaxique et les procédures d'interprétation du sens et aux difficultés d'isoler la première des secondes. Tout cela conduit à une conclusion pleine d'enseignements sur les tendances —et les erreurs— des théoriciens.

Quand nous essayons d'étudier le fonctionnement de l'esprit, nous avons fortement tendance à postuler des théories qui reposent sur une manipulation de symboles explicites par des règles explicites. Or cette hypothèse est une projection de la façon qu'ont les théoriciens, comme toute autre personne, de penser consciemment en termes de catégories bien définies. Mais quelles preuves avons-nous de l'existence d'une représentation mentale de règles grammaticales explicites? La réponse est que nous produisons et comprenons des phrases qui semblent suivre de telles règles, et que nos jugements sur ces phrases semblent aussi être guidés par ces règles. C'est tout. La théorie de la calculabilité montre qu'il y a un nombre infini de manières différentes d'effectuer un calcul. Il est donc tout à fait possible que nous n'ayons intégré ni les règles formelles de la grammaire, ni les règles formelles de la logique (p. 344).

Le même sens du complexe se retrouve dans le chapitre sur le sens, le plus intéressant sans doute du point de vue communicationnel. Que construisent les humains quand ils comprennent un énoncé et comment se déroule le processus de construction? Selon Johnson-Laird, la plupart des théories qui ont proposé des réponses à ces questions commettent l'erreur symbolique consistant à croire que le sens procède des relations entre symboles linguistiques. Ainsi des diverses théories des réseaux sémantiques qui supposent que les significations des mots (et phrases) peuvent être représentées à l'aide de réseaux comportant des liaisons avec des termes génériques (ou archi-lexèmes), des termes représentant des parties, des propriétés, etc. Ainsi des théories qui supposent que la

signification des mots peut être décomposée en atomes de sens primitifs et universels ou de celles qui, renonçant à cette dernière hypothèse, supposent que l'on peut inférer les conséquences nécessaires de certains mots au moyen de "postulats de sens" (comme "si x est une table, alors x a un plateau"). Toutes ces théories passent à côté d'un problème essentiel: celui de la relation des mots au monde.

Pourquoi tant de théoriciens sont-ils passés à côté de la relation entre le langage et le monde? Ils ont peut-être supposé qu'elle est indépendante du problème sur lequel ils se concentraient, celui des relations entre des locutions. Or ces deux types de relation ne sont pas indépendants. Mon opinion est qu'une théorie qui établit un rapport entre les mots et le monde relie peu à peu les mots les uns aux autres et rend donc superflue l'autre ensemble de théories qui accomplissent la seconde tâche (p. 354).

Cette théorie qui relie les mots au monde est celle des "modèles mentaux".

Dans la dernière partie, intitulée "Le conscient et l'inconscient", l'auteur s'attaque aux problèmes —ultimes— de la conscience et de l'inconscient, de la volonté, des sentiments. Nous y reviendrons dans la suite, mais nous voudrions d'abord présenter cette théorie des modèles mentaux à laquelle nous avons fait plusieurs fois allusion.

La *théorie des modèles mentaux* a été construite pour rendre compte d'un certain nombre d'opérations cognitives qui d'une part semblent fonctionner comme des déductions logiques, mais d'autre part sont à la fois plus rapides et moins rigoureuses que celles-ci. Il s'agit de comprendre comment l'esprit est capable de traiter un grand nombre d'informations, souvent de façon efficace et adaptée, sans passer par les déductions formelles qu'impliquerait une grammaire logique innée. Par exemple, il est possible

de faire une inférence valide en imaginant la situation décrite par les prémisses, en formulant une conclusion informative qui est vraie dans cette situation, et en voyant pour finir s'il y a moyen que la conclusion soit fausse. Imaginer une situation, c'est, selon moi, construire un "modèle mental" (...). Vous construisez donc un modèle fondé sur le sens des prémisses, et non sur leur forme syntaxique, et sur les connaissances générales déclenchées par leur interprétation. Ensuite, si possible, vous tirez du modèle une conclusion qui n'est pas explicite dans les prémisses et n'élimine pas l'information sémantique du modèle. Enfin vous cherchez d'autres modèles de prémisses qui contrediraient la conclusion. S'il n'y en a pas, c'est que la conclusion est valide (p. 241-242).

Selon Johnson-Laird, nous raisonnons peut-être selon trois méthodes: en utilisant des règles d'inférence formelles, des règles spécialisées et des modèles mentaux. Mais la théorie des modèles mentaux représente l'approche la plus économique du raisonnement humain. De plus elle s'applique aussi à la vision ou à la commande du mouvement et surtout à la compréhension linguistique. Elle "justifie à la fois la compétence logique —le potentiel de rationalité des humains— et les erreurs" qui peuvent résulter du "goulot d'étranglement" qu'est la mémoire de travail, lequel limite les possibilités de production ou de rétention des modèles qui peuvent contredire une conclusion. Les sujets, en effet, produisent toujours des conclusions qui ne sont compatibles qu'avec quelques uns des modèles possibles des prémisses (p. 246).

Du point de vue linguistique, la théorie des modèles mentaux montre que la question du sens —et de la compréhension du sens— des énoncés ne peut se réduire à une question intra-linguistique, mais implique d'abord une relation du langage avec le monde. Le "modèle mental", en effet, intervient entre les expressions linguistiques et le monde.

"Ma stratégie, écrit l'auteur, consiste donc à montrer comment les expressions linguistiques sont liées aux modèles (...) puis à montrer comment les modèles sont liés au monde et donc

comment on peut éviter l'erreur symbolique" (p. 354).

Partant de la distinction entre signification et référence —établie par Frege—, la sémantique formelle a montré que les phrases peuvent être interprétées en relation non avec le monde, mais avec un modèle. Mais la sémantique formelle dont les règles déterminent les référents des expressions n'est pas toujours applicable au langage humain. Du point de vue psychologique, le problème principal vient de ce qu'une phrase peut correspondre à une infinité d'images; l'assertion "la table est à droite du buffet" peut être vraie pour beaucoup de tables, buffets, et positions différentes. La théorie de Johnson-Laird suppose que la représentation mentale initiale d'une affirmation sert à construire un modèle, (qui sera révisé à la lumière d'informations ultérieures).

Cela a deux conséquences. En premier lieu, puisque la théorie des modèles vise à expliquer le rapport entre les mots et le monde, les relations entre les mots et les situations du monde

doivent donc reposer sur des éléments situés à un niveau inférieur aussi bien aux mots qu'aux scènes considérées. Ces éléments sémantiques doivent être indicibles — ils ne peuvent pas être exprimés dans le langage considéré, mais seulement au moyen d'une terminologie théorique spéciale (p. 359).

Par exemple, un énoncé comme "le lave-vaisselle est à droite du placard" doit conduire à la construction d'un modèle ayant même structure générale que celui que l'on constituerait en voyant ou imaginant la cuisine. Il s'ensuit que l'expression "à droite de" doit comporter les éléments primitifs relatifs à la profondeur et à l'horizontalité.

En second lieu, on ne construit qu'un seul modèle mental, un modèle unique et provisoire, mais qui, une fois construit, peut être révisé, au fil des énoncés successifs au moyen d'une procédure récursive. C'est ainsi que:

un seul modèle peut servir d'échantillon représentatif de l'ensemble potentiellement infini des modèles compatibles avec un énoncé, parce qu'il peut être révisé de façon à rester conforme à toute information cohérente ultérieure (p. 361).

Le modèle sert ainsi de support à des inférences valides qui, conformément à la théorie du raisonnement évoquée précédemment, ne nécessitent pas de règles d'inférence formelle.

La théorie des modèles mentaux apporte une contribution essentielle à la compréhension de la logique communicationnelle. Voici la formulation synthétique qu'en donne l'auteur:

Les humains perçoivent le monde et en construisent des modèles. Ils peuvent juger des affirmations sur le monde perceptible par référence à ces modèles et les manipuler afin de concevoir et de juger des affirmations sur des sujets abstraits. Ils peuvent aussi reproduire ces modèles dans le discours, c'est-à-dire produire des comportements symboliques — des expressions linguistiques — censés transmettre le modèle à quel-qu'un d'autre. De son côté, l'individu qui décode ces expressions linguistiques construit un modèle qui ressemble à l'état du monde que le locuteur a connu et voulu transmettre. Ce modèle correspond à une catégorie de situations, dont celle à laquelle la phrase devrait se référer. Il peut cependant être révisé en fonction d'informations ultérieures, de telle sorte que la représentation linguistique initiale de la phrase cerne leur signification — il indique l'ensemble des autres catégories de modèles compatibles avec sa vérité. Le langage nous permet donc de connaître le monde et d'en communiquer certaines notions abstraites (p. 368-369).

La question qui reste en suspens, dans les recherches de Johnson-Laird est celle de savoir jusqu'à si les "modèles mentaux" sont concrets et singuliers ou au contraire abstraits et généraux. Il est certain que nous pouvons avoir —pour reprendre l'expression de Kintsch et Van Dijk— un "modèle de situation", qui provient immédiatement de la vision d'un événement ou qui est suscité par le récit d'un événement singulier. Mais nous savons aussi que les événements ou les situations appartiennent à des catégories d'événements ou de situations (par exemple, l'événement "accident de voiture" ou la situation "à droite du

lave-vaisselle") et que donc nous construisons probablement des modèles mentaux de plus en plus schématiques, généraux et abstraits, susceptibles de donner une pré-interprétation plausible d'un grand nombre d'événements. La solution —élégante— de Johnson-Laird est que nous procédons par défaut (et que nous corrigeons ou révisons continuellement nos modèles mentaux en fonction des informations ultérieures). Mais cette solution ne permet pas de déterminer le degré de concrétude ou d'abstraction requis par les modèles mentaux: s'ils sont trop concrets, il s'adaptent difficilement à la multiplicité des situations. S'ils sont trop abstraits, ils ne peuvent plus jouer leur rôle d'aide à la compréhension car ils leur manque les éléments qui permettent de les accorder aux situations concrètes.

La conclusion de l'ouvrage, philosophique et épistémologique, est consacrée à la discussion des arguments pour et contre la science cognitive et constitue un plaidoyer en faveur de cette dernière et de l'intérêt des résultats obtenus en traitant l'esprit comme une machine.

L'ouvrage de Johnson-Laird contient en effet une théorie fortement articulée de la nature et du fonctionnement de l'esprit. L'architecture de l'esprit peut être comprise comme un vaste réseau de processeurs agissant en parallèle, la "conscience" consistant alors en un contrôleur de haut niveau, un "système d'exploitation" qui établit les objectifs des niveaux inférieurs. Cette conception hiérarchique implique qu'il y ait dissociation entre processus conscients et inconscients: alors que la conscience fonctionne de manière explicitement symbolique, les processus mentaux inconscients dépendent peut-être non de manipulations de symboles selon des règles, mais de représentations parallèles distribuées selon des modèles connexionnistes.

L'esprit conscient dépend du traitement en série de symboles qui ont une structure explicite, tandis que l'inconscient dépend du traitement parallèle de représentations symboliques distribuées. (p. 405)

Cela ne signifie pas du tout que l'auteur adopte une conception connexionniste du fonctionnement de l'esprit. D'une part, si le fonctionnement général de l'esprit est expliqué à partir l'idée de "réseaux agissant en parallèle", cela renvoie plutôt aux conceptions de Fodor sur la "modularité de l'esprit" ou à celles de Minsky sur la "société de l'esprit", et non pas du tout à l'idée d'un fonctionnement connexionniste généralisé. D'autre part, les processus spécifiquement intelligents renvoient à des traitements explicitement symboliques et donc à des modèles calculables classiques (quelle que soit par ailleurs leur complexité).

Ce dont il s'agit, en définitive, c'est bien de l'opposition fondamentale entre deux représentations de l'esprit et de son fonctionnement. La première est "symbolique", logique, formelle: on doit retrouver dans l'esprit des formes logiques que l'on découvre dans le raisonnement "correct", ou des équivalents de nos perceptions, souvenirs, hypothèses, etc. La seconde représentation est "connexionniste": il n'y a pas d'équivalent symbolique à nos représentations conscientes. Elle est "subsymbolique" (Smolensky). Johnson-Laird prétend combiner les deux conceptions dans un ensemble hiérarchisé, où le "conscient", ou du moins le symbolique explicite, occupe la position haute:

La théorie de l'architecture mentale que j'ai décrite (...) postulait que la conscience est un système d'exploitation situé au sommet d'une hiérarchie de processeurs. Le système reçoit des différents processeurs de la hiérarchie des messages qui représentent le monde et de son côté, il leur envoie des messages dans lesquels il leur communique ses plans. Les processeurs situés en dessous de la hiérarchie forment des modules qui peuvent utiliser des représentations distribuées, mais les communications avec le système d'exploitation demandent l'emploi de symboles ayant une structure explicite et un contenu propositionnel. (p. 396).

Cette thèse se révèle très bien dans la conception que l'ouvrage propose de la "conscience de soi", cette fonction qui dans la philosophie traditionnelle représente l'élément le plus irréductible à toute forme de traitement informatique. Pour Johnson-Laird, la conscience de soi proprement dite est sans doute une fonction du système d'exploitation mais nécessite un mode de traitement spécial. Il faut supposer que le système d'exploitation a accès à un modèle de lui-même qu'il peut utiliser de façon réflexive:

(...) le système d'exploitation demande la construction d'un modèle de ses opérations dont il se sert pour guider ses processus. Cette procédure réflexive peut être appliquée à ses résultats, afin que le système puisse construire un modèle de son emploi de tels modèles, et ainsi de suite, en une série de niveaux de méta-représentations de plus en plus élevés. (p. 382)

Ce qui est essentiel dans une telle représentation, c'est la subordination des modules "inférieurs" au module "supérieur", qui est le système d'exploitation. Il faut même aller plus loin. Les modules inférieurs sont importants dans la mesure où ils peuvent être insérés dans des processus spécifiquement cognitifs. Ainsi, par exemple, à l'origine des émotions, on doit trouver une évaluation cognitive des événements et l'origine des émotions complexes doit se situer dans les évaluations cognitives du système qui a accès à un modèle du moi. La fierté, le remords, le res-sentiment, la peur ou la haine de soi proviennent de l'évaluation de notre situation par rapport au modèle de nous-même.

Enfin, ce parti pris pour une théorie cognitiviste dure se reflète dans la conception que Johnson-Laird se fait des rapports de l'esprit et du monde. Il s'agit, en effet, d'une conception typiquement "représenta-tionniste" du monde. Il y a "un monde indépendant du langage et de la pensée", "il existe un monde physique dans l'espace et le temps" et "ce sont des processus mentaux qui permettent aux gens de percevoir ledit monde, d'avoir des pensées et des senti-ments, et d'avoir conscience d'eux-mêmes" et si ces hypothèses sont remises en cause "on entre dans le domaine de la métaphysique, qui se situe hors du champ de la science" (p. 411), car "pour autant que nous le sachions, toute théorie scientifique peut être mo-délisée sous forme de programme d'ordinateur" (p. 412). D'où l'idée optimiste que rien n'interdit de doter des ordinateurs "des mécanismes né-ces-saires pour avoir des motivations, des sensations et une cons-cience" (p. 412) (c'est-à-dire un rapport avec le monde). Cela signifie que Johnson-Laird adopte l'hypothèse de Turing selon laquelle "toute procédure fonctionnelle peut être réduite aux opérations d'un de ses automates élémentaires" et il en résulte que "quels que soient les changements apportés", l'auteur prédit "en toute sérénité que le concept de simulation sur ordinateur ne sera remplacé par aucun autre système" (p. 415) pour expliquer le fonctionnement de l'esprit humain.

Cette image des rapports entre esprit et monde explique, nous semble-t-il, pourquoi le "schéma de communication" que nous avons évoqué précédemment reste lui aussi tellement conventionnel. De ce point de vue, la théorie des modèles mentaux, qui aurait pu occuper une place centrale et constituer le pivot d'une théorie pragmatique de la communication, n'apparaît finalement que comme une manière commode de penser les relations de subordination entre d'une part les mécanismes distribués et informels de la perception immédiate des situations et de leur compréhension provisoire et d'autre part les mécanismes hautement programmables de la conscience et de l'intelligence informatique ou informatisable.

notes:

1. Philip N. JOHNSON-LAIRD, L'ordinateur et l'esprit, tr. J. Henry, Paris, Éditions Odile Jacob, 1994 (The Computer and the Mind, Fontana Press (Harper Collins), 1988, 1993 2ème éd.).

2. Johnson-Laird a élaboré et développé cette théorie dans *Mental Models. Towards a Cognitive science of Language, Inference, and Consciousness*, Cambridge, Cambridge University Press, 1983.