Revue Internationale de Systémique is published 5 times a year: March, May, July, September, December. Date of issue: March 1994.
Subscription price, per volume: Institutions US $ 190.
Second-class postage paid at Rahway, N.J. ISSN N° 0980-1472, USPS N° 007728.
U.S. Mailing Agent: Mercury Air-freight Intnl. Ltd., 2323 Randolph Ave., Avenel, NJ07001.
Published by Dunod, 15, rue Gossin, 92543 Montrouge Cedex France and Gauthier-Villars North America Inc., 875-81 Massachusetts Avenue, Cambridge, MA 02139, USA.
Postmaster. Please send all address corrections to: Dunod, c/o Mercury Air-freight Intnk. Ltd. 2323 Randolph Ave., Avenel, NJ 07001, USA.

REVUE INTERNATIONALE DE SYSTÉMIQUE
Vol. 8, N° 1, 1994, pp. 1 à 11

INTRODUCTION

INTELLIGENCE ARTIFICIELLE DISTRIBUÉE : MODÈLE OU MÉTAPHORE DES PHÉNOMÈNES SOCIAUX

Charles LENAY [1]

Cette publication a pour origine un travail sur les sciences cognitives et leur épistémologie mené depuis plusieurs années à l'Université de Technologie de Compiègne. Une de nos problématiques récurrentes concerne les fondements et limites de l'individualisme méthodologique qui caractérise l'approche dominante du cognitivisme : dans quelle mesure peut-on prétendre comprendre l'ensemble des phénomènes cognitifs en restant à l'intérieur de l'individu, que ce soit un sujet humain ou une machine d'Intelligence Artificielle (IA) ? C'est la recherche et l'évaluation des alternatives qui nous ont conduit à ce domaine émergent de l'Intelligence Artificielle Distribuée (IAD) et des Systèmes Multi-agents [1]. Leur principe directeur est que la résolution de problème, ou simplement la production de comportements cohérents dans un groupe, résultent d'une activité collective qui échappe en partie à chacun des agents du système [2]. Les perspectives sur la conception de systèmes cognitifs artificiels s'en trouvent profondément changées et le dialogue interdisciplinaire propre aux sciences cognitives est très largement renouvelé. Les Systèmes Multi-agents et l'IAD semblent pouvoir servir pour analyser et modéliser divers aspects des théories biologiques, sociologiques, historiques ou économiques, mais inversement, ils doivent tenir compte des résultats accumulés dans ces domaines. C'est ce dialogue que nous avons tenté d'encourager à Compiègne lors de séminaires interdisciplinaires en 1992, 1993 et 1994 [3]. Nous consacrons ce numéro à cette problématique théorique et épistémologique centrale : quelles sont les prétentions possibles ou acceptables des modélisations de l'IAD pour les sciences sociales ? Et inversement, quels sont les résultats des sciences humaines réellement employés pour la construction de ces systèmes artificiels ? [4]

Dans l'approche classique de l'Intelligence Artificielle et de la psychologie cognitive, la cognition est conçue comme un traitement individuel de

1. Université de Technologie de Compiègne, BP 649, 60206 Compiègne.

représentations symboliques : le système calcule une solution à partir des faits et des règles qui lui sont fournies. Ce calcul s'effectue en manipulant les symboles d'après leurs formes et non leurs significations. De tels systèmes de traitement de symboles physiques ne peuvent donc fonctionner correctement que si les symboles et les règles qui les lient sont une représentation suffisamment adéquate des choses et de leurs liaisons dans le monde. Le résultat est affiché sous forme d'une suite de symboles que l'utilisateur (ou l'interlocuteur) doit interpréter pour en saisir la signification. Au contraire, dans le cadre d'une approche multi-agents, le calcul réalisé par chaque agent n'a une efficacité dans le travail collectif que s'il est saisi par d'autres agents internes au système. L'intérêt et le rôle d'un agent est d'agir sur les autres agents et de percevoir leurs réactions. La conception d'un système multi-agents nécessite donc la mise en place d'un médium entre les agents qui soit susceptible de répercuter leurs actions. Chaque agent s'inscrit dans un environnement sur lequel il agit et dont il perçoit les modifications. Mais l'environnement de chaque agent est beaucoup plus que ces micro-mondes que l'on a pu modéliser pour l'IA et la robotique classique (comme pour le programme SHRDLU de Terry Winograd). Doté d'une structure, d'objets, et de relations modifiables par les actions, il contient aussi d'autres agents actifs.

Le premier article « La Kénétique : des systèmes multi-agents à une science de l'interaction » de Jacques Ferber met en place le cadre théorique et technique de l'IAD et des Systèmes Multi-Agents en général. La diversité de ces systèmes peut s'ordonner suivant une échelle de complexité des agents. A un extrême, on a des systèmes dont les agents ont des fonctionnalités très simples. Ils réagissent aux modifications locales de leur environnement en produisant de nouvelles modifications qui seront perçues par les agents suffisamment proches. On parle d'*agents réactifs*. En dépit de la simplicité des unités qui les composent, de tels systèmes peuvent donner lieu à des comportements d'ensemble extrêmement riches [5]. A l'opposé, on a des systèmes rassemblant des agents beaucoup plus complexes constitués comme des systèmes à base de connaissances (ou systèmes experts) de l'IA classique. Dans ce cas on parle *d'agents cognitifs*. C'est dans ce sens que furent développées les premières recherches de l'IAD. Ali Benmackhlouf, analysant les *Readings in Distributed Artificial Intelligence* [6] met au clair les termes fondamentaux et les problématiques principales de ce domaine. Un agent réalisant un traitement symbolique des connaissances devra représenter un monde contenant d'autres agents dont lui-même. Ses actions peuvent être de deux types. Ce sont soit des comportements qui modifient l'environnement des autres agents, soit des messages qui visent à modifier les connaissances des autres agents. A l'intérieur d'un système d'IAD, l'envoi d'un message est donc une action. C'est ce qui

explique l'emploi de la théorie des actes de langage dans de nombreuses modélisations de l'IAD. Marie-Pierre Gleizes, Pierre Glize et Sylvie Trouilhet dans leur article « Étude des lois de la conversation entre agents autonomes » présentent la façon dont les théories de la communication humaine (actes de langage, maximes de la conversation) peuvent être utilisées pour modéliser les interactions entre les agents d'un système d'IAD.

Si la signification d'une chaîne de symboles est l'effet qu'elle produit sur l'agent qui la reçoit (acte performatif), alors l'interprétation est un processus interne au système. Chaque agent est à la fois interprète des autres agents et objet de leurs interprétations.

Si, comme dans certains systèmes multi-experts, l'on réussit à éliminer toute ambiguïté interprétative (les symboles ont les mêmes rôles dans les calculs effectués par les différents agents), on sort du cadre des systèmes multi-agents proprement dit : toute distinction entre agents différents est purement artificielle et l'IAD ne devient qu'une méthode de programmation.

Cependant, même si un langage de programmation commun est employé pour construire les différents agents cognitifs, rien n'assure qu'aucun conflit ne surgira de leur collaboration. En effet, c'est un problème classique des systèmes à base de connaissances que celui du maintien de leur cohérence quand on essaie de leur apporter de nouvelles connaissances (c'est-à-dire ici des règles liant des symboles). Si, dans le meilleur des cas, le système expert a une cohérence interne qui résulte du lourd travail de sa conception, l'introduction d'une connaissance supplémentaire nécessite de vérifier à nouveaux frais l'ensemble de la cohérence du nouveau système. C'est pourquoi on ne peut réaliser ce rêve des pionniers de l'IA qui pensaient pouvoir cumuler indifféremment les connaissances les plus diverses. *A fortiori*, rien n'assure que deux systèmes à base de connaissances conçus indépendamment auront une interprétation semblable d'une même chaîne de symboles. Une des motivations premières du développement de l'IAD a donc été de construire des systèmes multi-experts en faisant l'économie d'un examen direct de la cohérence globale (qui nécessiterait en fait une reconstruction complète des différentes unités). Plutôt que de chercher à éliminer les contradictions, on construit un système de dialogue qui accepte l'existence de conflits. L'émission d'un message correspond à la mise en extériorité d'une suite de symboles puisque l'interprétation qui en sera faite échappe à l'agent émetteur. Un même message aura des effets différents, peut-être contradictoires, sur les différents agents. La problématique de la cohérence est remplacée par celle d'une gestion des conflits. Chaque agent est « encapsulé » dans une enveloppe de programme destinée à gérer ses relations et conflits avec les autres systèmes suivant des artifices plus ou moins ressemblant

à ceux que l'on peut décrire lors des discussions entre experts humains. Si la cible du message n'est pas définie par l'agent émetteur, il est inscrit dans le milieu. On parle de structure *blackboard* : comme un tableau noir dans une classe, c'est une mémoire commune en modification permanente, accessible à tout ou partie des agents, et où chacun vient prendre et modifier les messages qui l'intéresse.

Cependant, on doit se demander si tous ces anthropomorphismes sont justifiés par une tentative réelle de modélisation des phénomènes sociaux, ou s'ils ne sont que de simples métaphores pour aider à la conception et la programmation de systèmes artificiels sans rapport avec les systèmes naturels [7]. Dans son article « Intelligence artificiel distribuée : entre simulation et métaphore » Bruno Bachimont adopte une perspective très critique. Remontant aux sources des prétentions immenses de l'IA classique, il montre que l'IA Distribuée n'échappe pas aux réserves générales que l'on peut faire sur les fondements d'un traitement mécanique des connaissances.

On peut envisager deux façons d'appréhender les travaux sur les systèmes multi-agents. Soit, restant dans le cadre strict de l'IA, on cherche à construire des systèmes de *représentation* procédant par un calcul sur des représentations symboliques des connaissances, soit l'on utilise ces nouvelles possibilités de modélisation pour *simuler* des comportements sociaux.

Dans la première perspective, outre la signification interne des symboles manipulés, on doit aussi leur associer une interprétation externe. Le contenu sémantique d'un message émis par un agent ne se réduit pas à son effet sur un autre agent interne au système (son contenu « étroit » définit par son rôle fonctionnel). Il faut aussi lui associer un contenu « large » correspondant à l'interprétation que le concepteur ou l'utilisateur du système peut faire des symboles qui lui sont accessibles. Dès lors, si comme Bruno Bachimont, on refuse que les connaissances humaines puissent être assimilées à des représentations symboliques, il faut considérer que c'est de façon purement métaphorique que l'on dit qu'une connaissance est contenue dans un système d'IA, de même qu'il est métaphorique de dire qu'une connaissance est contenue dans un livre. Ces symboles n'ont de sens que s'ils sont interprétés par des sujets humains.

Si face à un système d'IA classique on peut avoir le sentiment d'une intelligence au travail, c'est seulement dans la mesure où le message émis en fin de calcul nous apporte une information relativement inattendue. Les immenses capacités combinatoires de la machine lui ont permis de construire un message non prévisible par notre intelligence humaine. Cependant tout le problème est la confiance que l'on peut accorder à ce résultat si nous ne pouvons en vérifier le mode de production. C'est là la question centrale de la validation pour les systèmes à base

de connaissances. De ce point de vue, les systèmes d'IAD ne font que multiplier les métaphores. Puisque qu'aucun des agents n'a une représentation explicite du traitement général du problème, le résultat final est d'autant plus inattendu et sa validation d'autant plus problématique. Les interactions entre agents (messages, conflits, etc.) ne sont que des métaphores des interactions entre connaissances humaines. Il ne dépend que de l'utilisateur de bien vouloir les interpréter ainsi. Nous remarquions plus haut que dans un système d'IAD, un même message peut être « interprété » différemment par les différents agents. Convenons de nommer « interprétations symboliques » ces interprétations internes au système qui sur la base de la forme des symboles et de la syntaxe des messages reçus, débouche sur un calcul déterminé, et « interprétation sémantique », l'interprétation externe que l'utilisateur peut faire du fonctionnement du système. On voit l'extraordinaire difficulté de la validation d'un système d'IAD. Outre les problèmes classiques de la validation de l'interprétation sémantique des structures formelles de chaque agent, il faut résoudre le problème de la validation de l'interprétation sémantique de la diversité des interprétations symboliques des différents agents, ainsi que de leurs modes de résolution des conflits !

Dans la seconde perspective, l'IAD est un outil qui permet de *simuler* le fonctionnement de théories sociologiques. Partant de quelques hypothèses sur les propriétés des agents naturels et sur leurs modes d'interaction, on demande à la modélisation de montrer si les phénomènes collectifs émergents sont bien semblables à ceux qui sont observés dans les systèmes naturels.

Il y a toujours de grandes difficultés à donner un contenu rationnel à la notion d'émergence. Le simple appel à des idées comme celles de nouveauté radicale des propriétés collectives ou d'irréductibilité des propriétés du tout à celle des parties, peut facilement verser dans une pensée magique. La modélisation par des Systèmes Multi-Agents me semble être une bonne approche pour comprendre et mettre en évidence de nombreux phénomènes d'émergence d'une façon concrète et précise. La caractéristique essentielle de ces systèmes est que les agents sont insérés dans un environnement qu'ils *ignorent* pour la plus grande part. Ces agents relativement autonomes n'ont qu'un accès limité aux modifications du milieu et des autres agents : les états internes d'un agent ne reflètent pas la diversité des états de l'ensemble du système. Les propriétés et fonctionnalités émergentes concernent la structure et la dynamique du système dans son ensemble (et, comme le remarque Jacques Ferber, ces propriétés collectives contraignent rétroactivement le comportement de chaque agent).

A l'extrême, avec les agents réactifs qui n'ont qu'un nombre très limité d'états internes possibles, les propriétés émergentes ne concernent que la structuration du milieu et la différenciation des rôles joués par les agents. Ce sont certainement là des systèmes idéaux pour une étude systématique des phénomènes d'émergence.

Avec les agents cognitifs complexes qui n'ont pas accès au calcul dans son ensemble on peut aussi parler d'émergence à propos du résultat finalement atteint. Cependant, plus profondément, ce qui a émergé, c'est une interprétation collective nouvelle des messages symboliques. En effet, l'état final a été atteint à travers une suite de confrontations entre les interprétations symboliques divergentes des messages échangés. Dans le meilleur des cas, la gestion de ces conflits a abouti à un résultat suffisamment stable : face à telle question le système répond de telle façon. On comprend mieux le difficile problème de validation auquel nous avons fait allusion : il tient au fait que plus rien ne nous guide dans l'interprétation sémantique de cette réponse puisque l'interprétation symboliques de la question par le système nous est maintenant inconnue (sauf à pouvoir entrer dans le détail des mécanismes de gestion des différents conflits).

Dans tous les cas, ce qui semble crucial pour que soit mise en évidence l'émergence de phénomènes collectifs, c'est la *limitation* des compétences des agents, c'est-à-dire une forme de modélisation de l'ignorance. Le caractère inattendu des phénomènes émergents résulte non pas directement de l'ignorance de l'observateur, mais d'une ignorance interne qu'il a fallu précisément définir. Ce n'est pas là une limitation locale de la performance mais une limitation constitutive de la compétence des agents. Cette limitation participe en tant que telle à la dynamique globale, et justifie l'emploi de la notion d'émergence. C'est à travers un partage de leurs ignorances mutuelles que les agents constituent un environnement commun stable et historiquement diversifié.

Dans son article « Emergence of Social Organizations in Non-Human Primates », Bernard Thierry présente les divers problèmes méthodologiques que rencontre l'éthologiste dans l'analyse des structures sociales animales, en particulier chez les primates. Quels sont les niveaux pertinents pour expliquer leurs structures socio-démographiques ? Si les caractères individuels et les interactions correspondent à des comportements observables, les autres niveaux d'explication comme celui des relations sociales stables entre individus définis (relation de dominance) ou surtout celui des structures collectives (réseau social ou structures profondes) ne sont peut-être que des abstractions de l'observateur pour décrire et classer les formes sociales sans pour cela les expliquer causalement. Un principe méthodologique individualiste semble s'imposer pour lequel les propriétés de la société ne doivent résulter que des interactions physiques entre individus et non de structures

abstraites transindividuelles. Cependant, admettre que les propriétés sociales émergent des interactions signifie qu'elles ne pourraient être directement trouvées dans les attributs des individus isolés. La contribution d'un individu à la collectivité est en partie conditionnée par ces propriétés sociales émergentes.

Mais comment expliquer ce retour des propriétés émergentes sur le comportement individuel ? L'hypothèse la plus tentante est d'admettre une représentation animale des structures sociales. La représentation chez un individu de sa relation avec tel autre individu défini déterminera tel comportement ritualisé. De même, à un niveau encore plus abstrait, les réseaux sociaux résultent des interactions entre ces relations. Si l'on veut quand même leur accorder une efficacité causale sur le comportement individuel, et dans la mesure où l'on ne voit pas des relations sociales interagir dans le monde physique, il est tentant d'admettre que l'animal puisse non seulement se représenter les relations sociales des autres congénères, mais aussi qu'il puisse les comparer et concevoir des relations entre types de relation.

Néanmoins, les quelques modélisations par des systèmes de type IAD qui ont été tentées, ont montré que la morphogénèse de phénomènes sociaux pouvait être atteinte avec des hypothèses beaucoup plus économiques sur les capacités individuelles (elles utilisent des agents dont les fonctionnalités sont limitées et qui ne sont pas dotés d'une faculté de représentation). La simulation permet de pallier la difficulté de déduire les propriétés émergentes d'une collectivité à partir de la connaissance des modes d'interactions interindividuels. On rencontre bien sûr le même type de difficulté en sociologie.

L'article de Véronique Havelange, « Sciences cognitives et tradition sociologique », reprend la question de l'individualisme méthodologique en analysant ses emplois en sociologie comme dans les sciences cognitives. En effet, l'individualisme méthodologique définit dans les sciences cognitives classiques (la cognition ne peut être qu'un processus interne à l'individu) s'articule aisément avec celui de la tradition sociologique (les structures sociales doivent s'expliquer en référence à des individus et non à un quelconque « sujet collectif »). Pourtant il semble devenu clair que l'organisation sociale, les lois du contrat social ou les prix du marché, ne pouvaient être le produit de la simple sommation des actions d'individus d'abord entièrement définis. Présentant les travaux de Jean-Pierre Dupuy, Véronique Havelange rappelle comment l'on a pu essayer de résoudre ces difficultés en complexifiant un individualisme méthodologique que l'on se refusait à abandonner. L'individu reste l'atome primitif du social mais l'on reconnaît que ses caractères sont en partie spécifiés par les états collectifs qu'il contribue circulairement à déterminer. Cette rétroaction du tout sur les parties s'explique par la

A l'extrême, avec les agents réactifs qui n'ont qu'un nombre très limité d'états internes possibles, les propriétés émergentes ne concernent que la structuration du milieu et la différenciation des rôles joués par les agents. Ce sont certainement là des systèmes idéaux pour une étude systématique des phénomènes d'émergence.

Avec les agents cognitifs complexes qui n'ont pas accès au calcul dans son ensemble on peut aussi parler d'émergence à propos du résultat finalement atteint. Cependant, plus profondément, ce qui a émergé, c'est une interprétation collective nouvelle des messages symboliques. En effet, l'état final a été atteint à travers une suite de confrontations entre les interprétations symboliques divergentes des messages échangés. Dans le meilleur des cas, la gestion de ces conflits a abouti à un résultat suffisamment stable : face à telle question le système répond de telle façon. On comprend mieux le difficile problème de validation auquel nous avons fait allusion : il tient au fait que plus rien ne nous guide dans l'interprétation sémantique de cette réponse puisque l'interprétation symboliques de la question par le système nous est maintenant inconnue (sauf à pouvoir entrer dans le détail des mécanismes de gestion des différents conflits).

Dans tous les cas, ce qui semble crucial pour que soit mise en évidence l'émergence de phénomènes collectifs, c'est la *limitation* des compétences des agents, c'est-à-dire une forme de modélisation de l'ignorance. Le caractère inattendu des phénomènes émergents résulte non pas directement de l'ignorance de l'observateur, mais d'une ignorance interne qu'il a fallu précisément définir. Ce n'est pas là une limitation locale de la performance mais une limitation constitutive de la compétence des agents. Cette limitation participe en tant que telle à la dynamique globale, et justifie l'emploi de la notion d'émergence. C'est à travers un partage de leurs ignorances mutuelles que les agents constituent un environnement commun stable et historiquement diversifié.

Dans son article « Emergence of Social Organizations in Non-Human Primates », Bernard Thierry présente les divers problèmes méthodologiques que rencontre l'éthologiste dans l'analyse des structures sociales animales, en particulier chez les primates. Quels sont les niveaux pertinents pour expliquer leurs structures socio-démographiques ? Si les caractères individuels et les interactions correspondent à des comportements observables, les autres niveaux d'explication comme celui des relations sociales stables entre individus définis (relation de dominance) ou surtout celui des structures collectives (réseau social ou structures profondes) ne sont peut-être que des abstractions de l'observateur pour décrire et classer les formes sociales sans pour cela les expliquer causalement. Un principe méthodologique individualiste semble s'imposer pour lequel les propriétés de la société ne doivent résulter que des interactions physiques entre individus et non de structures

abstraites transindividuelles. Cependant, admettre que les propriétés sociales émergent des interactions signifie qu'elles ne pourraient être directement trouvées dans les attributs des individus isolés. La contribution d'un individu à la collectivité est en partie conditionnée par ces propriétés sociales émergentes.

Mais comment expliquer ce retour des propriétés émergentes sur le comportement individuel ? L'hypothèse la plus tentante est d'admettre une représentation animale des structures sociales. La représentation chez un individu de sa relation avec tel autre individu défini déterminera tel comportement ritualisé. De même, à un niveau encore plus abstrait, les réseaux sociaux résultent des interactions entre ces relations. Si l'on veut quand même leur accorder une efficacité causale sur le comportement individuel, et dans la mesure où l'on ne voit pas des relations sociales interagir dans le monde physique, il est tentant d'admettre que l'animal puisse non seulement se représenter les relations sociales des autres congénères, mais aussi qu'il puisse les comparer et concevoir des relations entre types de relation.

Néanmoins, les quelques modélisations par des systèmes de type IAD qui ont été tentées, ont montré que la morphogénèse de phénomènes sociaux pouvait être atteinte avec des hypothèses beaucoup plus économiques sur les capacités individuelles (elles utilisent des agents dont les fonctionnalités sont limitées et qui ne sont pas dotés d'une faculté de représentation). La simulation permet de pallier la difficulté de déduire les propriétés émergentes d'une collectivité à partir de la connaissance des modes d'interactions interindividuels. On rencontre bien sûr le même type de difficulté en sociologie.

L'article de Véronique Havelange, « Sciences cognitives et tradition sociologique », reprend la question de l'individualisme méthodologique en analysant ses emplois en sociologie comme dans les sciences cognitives. En effet, l'individualisme méthodologique définit dans les sciences cognitives classiques (la cognition ne peut être qu'un processus interne à l'individu) s'articule aisément avec celui de la tradition sociologique (les structures sociales doivent s'expliquer en référence à des individus et non à un quelconque « sujet collectif »). Pourtant il semble devenu clair que l'organisation sociale, les lois du contrat social ou les prix du marché, ne pouvaient être le produit de la simple sommation des actions d'individus d'abord entièrement définis. Présentant les travaux de Jean-Pierre Dupuy, Véronique Havelange rappelle comment l'on a pu essayer de résoudre ces difficultés en complexifiant un individualisme méthodologique que l'on se refusait à abandonner. L'individu reste l'atome primitif du social mais l'on reconnaît que ses caractères sont en partie spécifiés par les états collectifs qu'il contribue circulairement à déterminer. Cette rétroaction du tout sur les parties s'explique par la

spécularité des interactions : chacun agit en fonction de l'autre qui lui-même circulairement agit en fonction du premier, et ceci pour l'ensemble de la collectivité. Le principe d'organisation est alors l'existence de « comportements propres émergents », ou « points fixes endogènes », qui correspondent à des états stables où l'on voit que chacun est en partie déterminé par les formes sociales à l'existence desquelles il participe. Cette conception du social formalisée en une théorie de l'individualisme complexe semble pouvoir être modélisée par des systèmes d'IAD. En effet, là aussi les structures collectives sont le produit des interactions entre agents, et simultanément, rétroactivement, les rôles et propriétés des différents agents sont contraints par ces propriétés émergentes.

Cependant, Véronique Havelange poursuit son analyse critique, et adoptant le point de vue d'une autre tradition sociologique représentée essentiellement par l'herméneutique, elle montre les limites de l'individualisme méthodologique, même sous cette forme complexe. Ce formalisme reste incapable de reconstituer le social qu'il a éliminé de ses prémisses. Ainsi, il ne semble pas pouvoir rendre compte ni de la diversification des intérêts et des conflits, ni de l'historicité des structures sociales.

L'approche multi-agent révèle peut-être ici toute sa richesse, du moins si l'on veut bien accepter les perspectives qu'elle permet d'ouvrir pour résoudre une partie de ces difficultés. Il faut insister sur l'importance qu'elle accorde à la modélisation d'un environnement concurremment à celle des agents. Il ne s'agit pas ici de poser que les structures sociales seraient déterminées de l'extérieur par la structure de l'environnement physique comme par une réalité externe prédéfinie et indépendante des agents, mais il faut bien reconnaître qu'en tant que médium des interactions c'est l'environnement qui est le lieu de la rétro-action de la collectivité sur les individus. Il est clair, du moins dans les modélisations des SMA que les points fixes de la dynamique collective sont attachés à des modifications stables du milieu. Nonobstant le fait que l'environnement de chaque agent soit sans cesse modifié par ses actions, il n'en joue pas moins en retour un rôle structurant sur ses comportements.

Dès lors que l'on est attentif aux traces laissées par l'activité sociale, c'est-à-dire à la mémoire technique (ce que Bernard Stiegler appelle les êtres inorganiques organisés [8]), il devient possible de doter les structures sociales d'une efficace qui ne passe pas seulement par leur représentation dans les individus, et ceci sans tomber dans l'illusion d'une intentionnalité du collectif. Les structures sociales n'existent que par et pour les agents, mais en même temps, c'est là un réalisme minimal, elles correspondent à des inscriptions matérielles qui perdurent plus ou moins dans le temps suivant une causalité qui leur est propre [9]. La langue, les

institutions, les mathématiques et les théories scientifiques ne sont pas seulement des représentations dans les têtes des hommes. Elles ont aussi donné lieu à des inscriptions matérielles (sons, textes, monuments, vêtements, outils, etc.) qui sont structurées et constituent l'environnement de chaque agent. On voit comment la prise en compte d'un milieu physique des actions permettrait de comprendre la diversité et l'historicité des phénomènes sociaux. Lieu d'une structuration qui perdure, modifié par les agents et leurs interactions, mais aussi toujours déjà là, il définit des contraintes préalables à ces actions, c'est le lieu d'une mémoire collective [10]. Rien n'assure que les interprétations de ces traces soient univoques, mais chaque agent entrant dans le jeu des interactions avec la communauté, devra agir sous ces contraintes.

Les tentatives traditionnelles d'explication ou de formalisation des phénomènes sociaux (théorie de la décision, théorie des jeux, logique multi-épistémique) sont maintenant confrontées à un double défi. D'une part, elles doivent montrer leur validité en concurrence avec les modélisations de l'IAD. Mais d'autre part, dans la mesure où l'IAD est une méthode de simulation des phénomènes sociaux (et éventuellement du traitement social des signes symboliques), elles peuvent aussi bien être mobilisées pour rendre compte des phénomènes sociaux *artificiels* de l'IAD.

La théorie de la décision devrait permettre de formaliser les actions des agents « rationnels » et de leur société. Cependant, elle est soumise à une rude critique par Samuel Guttenplan dans « Preference & Rationality ». Sur la base d'une variante étonnante du paradoxe de Condorcet développé par Kenneth Arrow, il examine les fondements de cette théorie du choix rationnel en particulier le principe d'un rangement ordinal des préférences (ou des utilités), et il montre que même le postulat d'une transitivité de la préférence n'est pas toujours possible. Dès lors, rien n'assure la cohérence globale d'un ensemble de choix individuels indépendants, qu'il s'agisse de plusieurs personnes ou même d'une suite de choix de la même personne. On retrouve là un phénomène d'émergence puisque le choix global est impossible ou incohérent alors qu'il a été construit à partir de choix individuels rationnels. La cohérence de l'état final d'un calcul distribué n'est jamais certaine, même si chaque agent semble cohérent avec lui-même. Cette conclusion très cruelle pour toute confiance naïve dans le travail coopératif des systèmes artificiels, s'applique aussi bien aux systèmes multi-agents naturels. La cohérence globale du collectif n'est jamais donnée d'avance, mais elle peut être recherchée, ce qui nécessite certainement la représentation individuelle des résultats du travail collectif. Même si la chose est définitivement impossible (ce qui est peut-être tant

mieux) il reste que c'est là un thème constitutif de notre histoire que la poursuite, aussi bien en science qu'en politique ou en philosophie, de systèmes complets et cohérents de la nature et de l'homme.

La théorie des jeux semble particulièrement pertinente pour tenter de modéliser et de comprendre la dynamique des interactions dans les systèmes multi-agents. Quoique même la théorie des jeux à deux joueurs ne soit pas encore complète, certains des résultats déjà acquis sont particulièrement intéressants. C'est me semble-t-il le cas de la solution originale au problème du dilemme du prisonnier proposée par Gilles le Cardinal et Jean-François Guyonnet dans « Comparaison de trois approches stratégiques de la coopération ». En effet, l'idée suivie est de demander à chaque joueur de se déterminer non seulement en fonction de son bénéfice personnel espéré, mais aussi en fonction du bénéfice obtenu par l'adversaire. La référence pour le choix de chaque agent à cet état particulier de l'environnement permet de définir très simplement les comportements de peur, de tentation ou d'attrait pour la coopération.

Dans « The logical way of describing societies », Jacques Dubucs présente les fascinantes possibilités de formalisation des logiques multi-épistémiques. Elles semblent permettre de décrire les différents domaines de connaissance d'une multiplicité d'agents, ainsi que l'emboîtement de leurs représentations (A sait que B sait que A sait que etc.). Cependant, ce type de formalisme conduit à poser que chaque agent épistémique est doté d'une omniscience logique : entièrement cohérent avec lui-même, il a immédiatement accès à toutes les vérités logiques ! De plus on admet le plus souvent un axiome dit « axiome d'introspection négative » pour lequel chaque agent connaît tout ce qu'il ignore. Ce type d'idéalisation tout à fait irréaliste montre peut-être là les limites de la logique formelle pour décrire des agents dont les actions sont réalisées dans un monde qui leur est constitutivement inconnu, et dont les conséquences leurs sont irréductiblement inattendues.

Cependant, les modélisations de l'IAD rencontrent une difficulté équivalente. L'ignorance des agents est définie par une limitation de l'accessibilité à l'environnement. Mais une telle approche touche là ses propres limites. On se donne au départ une extériorité. Même si l'on restreint au maximum les attributs de l'environnement préalable aux interactions des agents, il reste que les limites de chaque agent ont été spécifiées du point de vue de cette extériorité. La modélisation multi-agent doit nécessairement poser des hypothèses qui outrepassent les limites que l'on voudrait modéliser. Il n'y a ni formalisation, ni modélisation informatique complète possible de l'ignorance humaine, de la finitude intrinsèque de notre pensée. Gageons cependant que les constructions des systèmes multi-agents permettront d'enrichir notre connaissance de nos ignorances et de mieux comprendre comment leur dynamique collective est constitutive du monde que nous partageons.

Notes et références

1. J. ERCEAU, J. FERBER, L'intelligence Artificielle Distribuée, *La Recherche*, 1991, n°233, p. 750-758.

Y. DEMAZEAU, J.-P. MULLER, *Decentralized artificial intelligence, Proceedings of the 1st modelling autonomous agents and multi-agents worlds* MAAMAW, Cambridge, 1989, North Holland, 1990, *Second Proceedings MAAMAV,* North Holland, 1991, E. WERNER, Y. DEMAZEAU, *Third Proceedings MAAMAW,* North Holland, 1992.

2. Ce domaine de recherche est en développement rapide. En France, on notera par exemple la création d'un groupe de travail à l'AFCET en 1992 par Jean Erceau, les premières journées Francophones Intelligence Artificielle Distribuée et Systèmes Multi-Agents organisées à Toulouse par l'AFCET à l'IRIT en 1993, ou l'école internationale d'informatique de l'AFCET organisée à Neuchatel au mois d'août 1993.

3. De plus, nos collègues économistes de l'UTC, organisant eux aussi durant les mêmes périodes un séminaire sur des thèmes parallèles, nous avons pu programmer des conférences et tables rondes communes.

4. La publication des actes de nos séminaires aurait été trop lourde. Dans l'esprit de dialogue de ces journées il n'était pas demandé de communication écrite aux intervenants. L'espace de ce numéro n'a pas permis de retenir les nombreuses autres communications passionnantes de ces séminaires : près de soixante interventions sur deux semaines !

5. Un dossier dévolu à ces systèmes réactifs sera prochainement publié dans *Intellectica*, la revue de l'Association pour la Recherche Cognitive.

6. A. H. BOND, L. GASSER (Eds), Morgan Kaufmann, 1988.

7. Le terme « métaphore » est employé ici dans le sens restrictif et négatif de la substitution analogique d'une chose à une autre. Cependant cette notion mériterait d'être analysée pour elle-même. La simple opposition du littéral et du figuratif semble plutôt obscurcir le rôle réellement constituant de cette figure rhétorique dans le fonctionnement du langage.

8. B. STIEGLER, *La technique et le temps t.1, La faute d'Epimethée*, Galilée, 1994.

9. Cette forme de principe d'inertie ou de permanence de l'objet peut bien sûr être modélisée de diverses façons et, pourquoi pas, intégrer une dégradation entropique plus ou moins rapide de toute structure de l'environnement.

10. Le séminaire que nous organisons à Compiègne cette année 1994 avec Véronique Havelange et Bernard Stiegler propose une rencontre interdisciplinaire autour de cette notion fascinante de « mémoire collective ».