Entre réductionnisme méthodologique et stratégie intentionnelle, l'éthologie, un modèle alternatif pour l'I.A.D. ?

Alexis Drogoul et Anne Collinot

Equipe MIRIAD – Thème OASIS - LIP6
UNIVERSITE PARIS 6
4, Place Jussieu 75232 PARIS CEDEX 05
e-mail: {Alexis.Drogoul, Anne.Collinot}@lip6.fr
http://www-poleia.lip6.fr/~{drogoul,collinot}

RESUME. L'éthologie est à la mode en I.A. et en I.A.D. Mais, si l'on regarde ce qui intéresse les chercheurs de ces disciplines dans l'éthologie, on s'aperçoit que ce sont au mieux des modèles de comportement, au pire des exemples ou des anecdotes tirés de l'observation des animaux. Or, ces modèles n'ont pas été conçus n'importe comment, et ces observations pas faites au hasard : les adopter ou les utiliser devraient inciter les informaticiens à infléchir leur méthodologie de recherche en prenant en compte la caractéristique principale de l'éthologie, qui est d'arriver à concilier des discours explicatifs réductionnistes *et* intentionnalistes. Nous présentons dans cet article quelques caractéristiques de cette démarche, et montrons brièvement en quoi elle peut être intéressante pour l'IAD, à la fois pour la conception d'agents et la conception d'organisations.

1. Introduction: l'éthologie

On peut dater l'apparition de l'éthologie moderne des travaux de Konrad Lorenz et Nikolaas Tinbergen [Lorenz 1984; Tinbergen 1971] , réalisé principalement entre la fin des années 1930 et le début des années 1950. Cependant, cette science n'a évidemment pas été créée *ex nihilo*, et de nombreux concepts clés sont apparus dans des travaux antérieurs, ceux des psychologues comparatistes de la fin du XIXème siècle [Loeb 1900; Morgan 1894] ou des béhavioristes de la première moitié du XXème siècle [Skinner 1938; Watson 1919] . Par extension, et bien que ce ne soit pas tout à fait exact sur le plan épistémologique, ces travaux sont considérés comme faisant partie de l'éthologie. L'éthologie est souvent succinctement définie comme la science dont l'objet serait l'étude des comportements animaux dans leur milieu naturel. Cette définition factuelle, si elle se révèle exacte quant à la qualification de la démarche expérimentale mise en œuvre par les éthologues, sous-estime cependant le formidable potentiel théorique que constituent les milliers de travaux qu'ils ont effectués depuis les cinquante dernières années, qui ont consisté à décrire, classifier, mesurer, comparer, expliquer et comprendre l'ensemble des comportements individuels et collectifs effectués par les êtres vivants peuplant la planète (y compris l'être humain). En biologie animale, effectuer ce travail peut être réalisé en utilisant quatre schémas explicatifs distincts :

— le schéma causal, qui présuppose la connaissance des facteurs de causalité endogènes (physiologie, rôle des hormones, etc.) ou exogènes (stimuli , interactions, etc.). On explique *comment* le comportement se déclenche, quelles en sont les *causes*.

— le schéma fonctionnel, qui tente de comprendre la *fonction* du comportement, c'est-à-dire sa conséquence adaptative[1]. On explique *pourquoi* il se déclenche, dans quel but, et quelle est sa fonction.

— le schéma historique final, qui se base sur l'origine évolutive du comportement. On explique le *comment* et le *pourquoi* du comportement, non plus par rapport à une instance spécifique, mais par comparaison avec des comportements similaires apparaissant dans d'autres espèces, et sur une échelle de temps supérieure.

— le schéma historique proximal, qui s'attache à son ontogénie, c'est-à-dire l'étude de son élaboration et de son perfectionnement durant la vie de l'animal. On explique *de quelle manière* un comportement est acquis, ou de quelle manière un comportement inné est modulé, sur une échelle de temps assez courte.

La caractéristique de l'éthologie, par rapport à cette classification, est qu'elle va tenter d'intégrer des explications causales, fonctionnelles et ontogéniques pour

---

[1] L'emploi du terme *fonction* en biologie est différent de l'emploi commun, puisqu'il fait implicitement référence au gain adaptatif potentiel que procure un comportement, et non uniquement au mécanisme permettant d'atteindre un but particulier.

expliquer un comportement, en essayant donc de répondre simultanément au *comment* et au *pourquoi* de son déclenchement. Le schéma historique final est surtout destiné (quand il est disponible) à fournir des hypothèses sur la fonction du comportement[2], et forme le plus souvent la toile de fond de l'explication[3].

La situation de l'explication éthologique est donc assez singulière : détaillée pour l'occurence d'un comportement, elle doit être suffisamment générale pour autoriser sa comparaison avec d'autres; objective (ou, en tout cas, neutre) dans l'observation, elle doit être prédictive pour permettre d'orienter les observations ultérieures; réductionniste dans son essence, elle doit être compatible avec des explications intentionnelles (l'homme a aussi un comportement). C'est une position en tout cas unique par rapport à la biologie en général, et par rapport à la psychologie.

Cet article a été conçu de façon à brièvement présenter l'intérêt que peut avoir cette démarche pour l'Intelligence Artificielle Distribuée, et, plus généralement, pour les domaines traitant de cognition artificielle (Intelligence Artificielle et Vie Artificielle, en particulier). Il est organisé en deux parties, la première traitant de la méthodologie et des modèles comportementaux, la seconde des modèles sociaux :

— Méthodologie et modèles comportementaux : nous présentons quelques éléments de la méthodologie induite par l'objet de recherche de l'éthologie et l'intérêt qu'elle peut avoir pour les sciences de l'artificiel, notamment au travers du développement de modèles comportementaux, qui ont déjà fortement influencés la réalisation d'agents artificiels.

— Modèles sociaux : le point précédent incite naturellement à considérer avec attention l'analyse que fait l'éthologie des comportements grégaires des animaux, qu'ils soient considérés comme sociaux ou non.

2. Méthodologie: une dialectique réductionnisme/intentionnalisme

La méthodologie suivie par les éthologues dans la conduite de leurs recherches est relativement diffuse et n'a jamais vraiment fait l'objet, si l'on excepte [Richard 84,85] et [Vauclair 1992], d'une qualification très précise. Comme toute méthodologie utilisée en biologie, si l'on met de côté ses aspects expérimentaux, elle se caractérise essentiellement par un travail particulier de *reverse engineering*. Confronté au comportement d'animaux qui ont été *conçus* par la sélection naturelle[4], l'éthologue va tenter d'en expliciter les causes, exactement comme un ingénieur de Peugeot pourrait analyser le comportement routier de la dernière Renault. Expliciter

[2] Ce schéma est surtout utilisé par les éthologues se réclamant de la sociobiologie, ou d'une de ses branches (voir, par exemple, pour un auteur français, [Jaisson 1993]).

[3] Dans le sens où un discours explicatif, quelqu'il soit, sera contraint par la théorie évolutionnaire. Comme le dit clairement [Dawkins 1976], tous les discours hypothétiques doivent, tôt ou tard, être traduisibles en langage adptationniste.

[4] C'est-à-dire en faisant référence à la sélection naturelle comme *processus de conception.* Il peut être utile de se référer à [Dennett 1995] pour comprendre le sens de cette implication.

un comportement nécessite bien entendu de disposer d'un certain nombre de données physiologiques et écologiques (comme l'ingénieur qui dispose du travail effectué par les personnes qui ont étudié la voiture en atelier, et qui sait également dans quelles conditions il conduit la voiture). L'éthologie s'appuie donc sur toutes les autres disciplines de la biologie ou de l'écologie, les premières lui fournissant les bases à partir desquelles construire des hypothèses sur le comportement ou sur la communication, les secondes les hypothèses sur l'environnement sensoriel dans lequel est plongé l'animal.

A partir de ces informations et des observations qu'il aura réalisées soit en laboratoire soit en milieu naturel, l'éthologue construit peu à peu un modèle de comportement, c'est-à-dire un schéma explicatif de la relation entre des causes proximales endogènes et exogènes et une manifestation comportementale particulière. Cependant, connaître les possibilités et le contexte ne permet pas toujours de trouver la bonne articulation entre une cause proximale et un comportement : l'éthologue a alors recours à des hypothèses, qu'il ne construit pas de n'importe quelle façon. C'est dans ce processus de construction d'hypothèses que l'on peut appréhender l'originalité de l'éthologie, aussi bien par rapport à la psychologie que par rapport à la biologie. On peut en effet en retenir deux points essentiels, qui peuvent apparaître contradictoire, mais qui représentent justement un discours explicatif dialectique extrêmement riche:

— l'usage d'un "bon" réductionnisme, essentiellement basé sur le *canon de Morgan*, pour la caractérisation causale et ontogénique du comportement,

— l'adoption d'une stratégie intentionnelle pour sa caractérisation fonctionnelle, c'est-à-dire l'explicitation des *raisons* du comportement.

2.1 *Le canon de Morgan : un "bon" réductionnisme*

Dans son *Introduction à une psychologie comparative*, datée de 1894, Lloyd Morgan énonça ce qui allait être connu par la suite comme le *Canon de Morgan* : «*En aucun cas, nous ne pouvons interpréter une action comme la conséquence d'un exercice ou d'une faculté psychique plus haute, si elle peut être interprétée comme l'aboutissement d'une faculté qui est située plus bas dans l'échelle psychologique* ». Cette prise de position, qui se démarquait clairement des travaux contemporains (marqués par l'idéalisme), allait avoir une influence considérable par la suite, en permettant l'émergence de courants de recherche dédiés au "psychisme animal" débarrassé de tout "psychologisme" : le béhaviorisme, dans une version intégriste, et l'éthologie, dans une version pragmatique.

La version intégriste, fort justement appelée par E. Sober béhaviorisme méthodologique, est définie comme "*la thèse selon laquelle il n'est pas nécessaire de faire appel à des états internes pour expliquer le comportement*" [Sober 1985]. Elle a donné naissance à un ensemble de travaux extrêmement riches, marqués sans doute par une idéologie un peu trop rigide, mais dont on a tendance à oublier, à l'heure actuelle, qu'ils ont mobilisés l'ensemble de la recherche en psychologie pendant à peu près cinquante ans avant d'être balayés en quelques années par la montée en puissance de la psychologie cognitive (et de sa comparse, l'I.A.).

La version pragmatique (inexactement citée comme étant la version originelle dans [Vauclair 1992]) revient par contre à affirmer «*qu'il n'est pas nécessaire de faire appel à des structures psychologiques d'ordre supérieur (...) quand des systèmes explicatifs plus simples (réflexes, par exemple) peuvent rendre compte de façon adéquate d'un comportement donné*». L'important dans cette version est qu'elle fonde réellement une démarche méthodologique, certes réductionniste, mais ne niant pas *a priori* l'existence de structures psychologiques d'ordre élevé ni d'états internes. Elle privilégie ainsi la recherche de systèmes explicatifs les plus simples possibles, mais laisse ouverte, une fois épuisée la voie réductionniste, la possibilité d'adopter un point de vue intentionnaliste. C'est ce qui permet, entre autres, à l'éthologie de concilier des points de vue extrêmement divers sur la cognition, et d'apparaître comme un pont entre béhaviorisme et psychologie cognitive.

2.2 *Une posture intentionnelle*

Comme nous l'avons dit plus haut, un des buts de l'éthologie est d'arriver à fournir une caractérisation fonctionnelle du comportement, en liaison avec les mécanismes causaux de son déclenchement et, éventuellement, son histoire. Or, toute caractérisation fonctionnelle repose sur une double *interprétation* :

— l'interprétation du gain adaptatif (et donc, reproductif) obtenu par l'animal quand il met en œuvre ce comportement. Le plus souvent, celle-ci n'est rendue possible que par comparaison avec des espèces ne possédant pas ce comportement[5].

— l'interprétation de la fonction (au sens commun du terme) que remplit le comportement, c'est-à-dire du problème qu'il résout ou du but qu'il permet d'atteindre. C'est là un point délicat, car il repose sur l'objectivité supposée de l'observateur, et sur sa capacité à prendre en compte même le plus infime détail du produit de ce comportement.

Cette caractérisation fonctionnelle, dans la perspective de la théorie évolutionniste, n'a évidemment de sens que si elle est généralisable à d'autres espèces. Et elle ne devient généralisable qu'à partir du moment où elle est capable de s'abstraire des mécanismes proximaux qui produisent les comportements. Ainsi, la fonction de nidification observée chez toutes les espèces d'oiseaux présente-t-elle des formes extrêmement variées, depuis le parasitisme du coucou jusqu'aux constructions complexes et hautes en couleurs de l'oiseau jardiner, en passant par les constructions collectives du moineau républicain, etc. Chacune de ces formes met en jeu des comportements totalement différents d'une espèce à l'autre, qui, pourtant, peuvent être rassemblés par l'éthologue sous un même vocable clairement *intentionnel*, c'est-à-dire prêtant aux différents oiseaux étudiés une même

---

[5] d'où l'importance de considérer des catégories généralisables à un ensemble d'espèces.

intentionnalité. Or, où réside cette intentionnalité ? Il est plus que probable, comme le souligne [McFarland 1990, p. 396] que "*[cette fonction] s'est développée uniquement parce qu'[elle] accroît la fonction de reproduction de ses porteurs; cela n'implique aucunement que l'évolution est dirigée à cette fin, ou que le comportement de celui qui construit un nid lui-même est dirigé vers ce but*". L'intentionnalité est donc ici clairement attribuée par l'observateur dans un but *heuristique* : faciliter la description et la comparaison de comportements apparemment tendus vers un même but et assurant un même succés reproductif[6].

On se retrouve ainsi devant un paradoxe méthodologique: étant donnés des comportements décrits de la façon la plus concise possible (selon le Canon de Morgan), "*ce qui nous intéresse est de demander quels gains en perspicacité, en pouvoir prédictif, en généralisation, pourraient être augmentés si nous adoptions une hypothèse d'ordre supérieur qui prendrait le risque de donner une caractérisation fonctionnelle*" [Dennett 1990, p. 321].

Mais, contrairement à ce qu'affirme par exemple [Engel 1994][7], le paradoxe n'est qu'apparent, et c'est au contraire une démarche dialectique d'une très grande richesse qui naît de l'opposition de surface entre les deux démarches. Nous avons en effet affaire ici à deux niveaux d'explication distincts (suivant en cela les schémas présentés plus haut) : une explication causale, proximale, qui sera susceptible d'être guidée par le Canon de Morgan; une explication qui tente de donner des *raisons* au comportement, par exemple en relation avec des comportements similaires dans d'autres espèces, et qui sera, elle, guidée par la posture intentionnelle. Comme le souligne ainsi [Davies 92], "*Une explication purement causale de l'acte n'est pas en compétition avec une explication qui donne des raisons*", mais elle en est au contraire complémentaire (et elle lui est sans doute nécessaire).

Ainsi, par exemple, l'explication intentionnelle d'un comportement de "tromperie" (chez les singes, les oiseaux, les insectes, etc.) peut très bien s'accommoder d'explications causales en terme de simples tropismes[8] comme d'explications supposant l'existence d'états internes (croyances, désirs, etc.) chez ces animaux[9]. Mais l'intérêt de regrouper ces comportements sous un même vocable intentionnel est évident : comme le montre fort bien [Dumouchel 1992], "*[la posture intentionnelle en éthologie] permet de tester divers degrés d'hypothèse pour un*

---

[6] Evidemment, il est probable que peu d'animaux, sinon aucun, ne possède la forme d'intentionnalité qui leur est prêtée de cette façon. Mais les tenants de la posture intentionnelle sont en général opposés à l'idée même de l'existence d'une intentionnalité *intrinsèque* : pour eux, les raisons d'un comportement, qu'elles soient *dans* l'animal ou *dans* la nature (les "raisons flottantes" de Dennett) peuvent et doivent être décrites de la même manière.

[7] qui oppose le Canon de Morgan à la posture intentionnelle, comme démarches méthodologiques, en éthologie (voir notamment pp.103-105)

[8] attractions / répulsions vis-à-vis d'objets ou sujets de l'environnement. L'explication généralisée de tout comportement en termes de tropismes, dùe à [Loeb 1900], est considérée comme l'une des premières explications "pré-éthologiques" non-psychologiques.

[9] l'attribution de facultés sera bien entendu doublement dépendante, à la fois des données empiriques disponibles, et des contraintes phylogénétiques. Il sera ainsi plus difficile, au regard de nos connaissances, d'attribuer des croyances à une fourmi qu'à un primate.

même comportement, de la morne explication béhavioriste d'ordre zéro à une explication intentionnelle romantique de quatrième ordre. [Elle] doit être jugée à sa capacité de suggérer des hypothèses intéressantes et de mettre en lumière des ressemblances jusqu'à ce jour inaperçues entre des comportements différents", c'est-à-dire entre des explications proximales *a priori* sans rapport.

C'est là sans doute que se situe le grand intérêt de cette démarche méthodologique : elle fournit un cadre particulièrement élégant permettant de confronter des thèses aussi diverses que le behaviorisme et les différentes formes de psychologie expérimentale pour expliquer un comportement [Griffin 1982]. Elle est bien sûr orientée par le Canon de Morgan dans le sens de la moindre complexité psychologique, mais, du fait qu'elle attribue à des animaux simples une forme d'intentionnalité (intrinsèque ou dérivée, peu importe), peut très bien s'accomoder, dans le cas d'animaux plus complexes, d'explications supposant l'existence d'une intentionnalité "intrinsèque" (comme c'est le cas pour les primates).

2.3 *Leçons*

C'est donc entre une certaine forme de réductionnisme et une nécessaire stratégie interprétative que se place la démarche méthodologique de l'éthologie. Que peut-elle nous apprendre ?

— Tout d'abord que l'on peut avoir une démarche *réductionniste* qui n'est pas forcément *réductrice*. Des deux versions du Canon de Morgan, la béhavioriste l'est, qui se ferme sciemment des ouvertures possibles vers la psychologie. La version éthologique est une forme de "bon" réductionnisme, au sens où l'entend [Dennett 1995], c'est-à-dire une démarche constructiviste : c'est à partir de briques de base les plus réduites possibles que l'on construit des briques un peu plus "cognitives"[10].

— Ensuite qu'une démarche réductionniste peut très bien s'accompagner d'une démarche intentionnaliste : non seulement il n'est pas stupide d'attribuer des états, ou une intentionnalité, à un système dont on sait pertinemment qu'il ne les possède pas en propre, mais il peut s'agir d'une démarche heuristique, permettant, pour une même catégorie de problèmes, de considérer différentes solutions technologiques.

C'est là, nous semble-t-il, que l'approche éthologique du problème de la cognition animale [Pearce 1987] peut être fondamentale pour les domaines qui souhaitent concevoir une forme de cognition artificielle : l'IA, bien sûr, mais également, comme nous le verrons plus loin, l'IAD[11]. Certes, l'éthologie s'attache à comprendre et à expliquer des systèmes existants, alors que l'IA ou l'IAD tentent d'en construire : à une démarche de *reverse engineering* semble donc s'opposer une

[10] Voir [Drogoul 1993], pour une explicitation du passage réductionnisme/constructivisme.

[11] En réalité, le problème n'est pas foncièrement différent, que la cognition considérée soit individuelle ou collective.

démarche plus classique d'ingénierie. Mais, en définitive, quelle est la différence ? Un éthologue passe autant de temps à concevoir des hypothèses ou à construire des modèles qu'à observer des animaux, et un chercheur en IA passe sans doute autant de temps à analyser (empiriquement ou théoriquement) les systèmes qu'il produit qu'à les concevoir. Comme le souligne Jean-Pierre Müller (comm. pers.), il n'y a pas de différence visible entre une démarche méthodologique *d'ingénierie* et une démarche de *reverse engineering*.

Distinguer les causes proximales qui déclenchent les comportements des raisons qui les motivent, en tentant, pour les premières, d'aller le plus loin possible dans la réduction et, pour les secondes, de les imaginer éventuellement "en-dehors" de l'agent considéré est une démarche qui peut s'exprimer aussi bien en éthologie qu'en IA. Elle apparaît même comme nécessaire dans le sens où, au-delà d'une implémentation particulière qui fournira le comportement adéquat, l'intentionnalité d'un système d'IA ou d'un agent n'est certainement pas une intentionnalité intrinsèque, mais au mieux une intentionnalité dérivée de celle de son concepteur.

Une autre caractéristique de la démarche éthologique est qu'elle prend le risque de développer une science du comportement privée de tout postulat de rationalité (sans lequel la psychologie ne peut fonctionner). Pour une IA et une IAD encore largement dépendantes de ce postulat (cf. la notion d'agent rationnel) c'est une démarche qui ne manque pas d'attraits, et qui permettrait, en tout état de cause, de ne pas laisser ce terrain exclusivement à la Vie Artificielle (qui, elle, repose un peu trop sur le postulat d'optimalité de la sélection naturelle) [Belew 1991].

2.4. *Hypothèses comportementales individuelles*

Les deux aspects méthodologiques exposés plus haut ont tout d'abord trouvé leur traduction dans la mise au point de *modèles comportementaux* génériques. C'est une des raisons pour lesquelles l'éthologie a suscité un grand intérêt en Vie Artificielle et en Intelligence Artificielle ces dernières années (voir [Meyer et Guillot 1990] pour une intéressante revue), puisque les modèles considérés comme classiques, celui de K. Lorenz et celui de N. Tinbergen, ont été presqu'intégralement repris dans un très grand nombre de travaux d'IA[12].

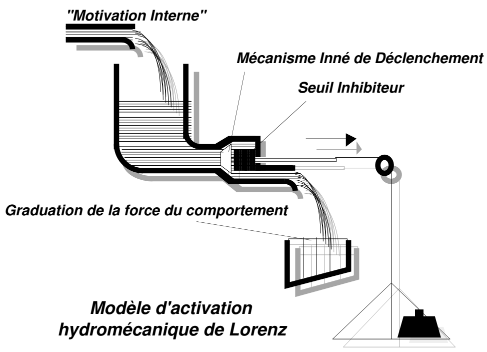
Contrairement aux modèles développés en psychologie cognitive, qui reposent de manière explicite sur l'existence d'une intentionnalité *intrinsèque*, donc sur une hypothèse de rationnalité, les modèles développés en éthologie reposent sur l'hypothèse implicite qu'une bonne partie de l'intentionnalité réside en réalité dans l'interprétation de l'observateur : la complexité du comportement apparent provient de la complexité de l'environnement dans lequel il est effectué (environnement physique, ou social, d'ailleurs). Il n'est donc pas forcément nécessaire de partir de cette hypothèse de rationnalité pour construire un agent : on peut se servir du Canon

---

[12] Nous ne prendrons que ces deux exemples dans cet article afin de ne pas alourdir l'exposé, mais il faut savoir que la littérature éthologique abonde en modèles de comportement descriptifs ou explicatifs (voir [Griffin 1982] ou [Guillot 1986]).

de Morgan pour l'éliminer en première instance (quitte à la reconsidérer par la suite si cela s'avère nécessaire). Il n'est cependant pas envisageable non plus de nier l'existence, dans tout comportement, même effectué par le plus simple des animaux, d'une forme de motivation indépendante de toute cause exogène[13] ou de capacités de réaction différentielle[14] à une même situation. Cette plasticité parfois surprenante amène tout naturellement à tenir un discours intentionnel sous la forme de désirs, croyances, buts, etc. c'est-à-dire à projeter ce que nous pensons connaître de nos états mentaux dans le modèle[15].

Un bon exemple de cette démarche réductionniste sur le mécanisme et intentionnelle dans son vocabulaire est fourni par le modèle classique de K. Lorenz [Lorenz 1984] [16]: dans ce modèle (voir figure), les catégories intentionnelles correspondant à la motivation, à la réceptivité, etc. sont en effet traduites sous la forme de mécanismes hydrauliques ou mécaniques qui ne reposent sur aucune hypothèse de rationnalité. L'usage d'un vocabulaire intentionnel pour décrire ces mécanismes, loin de le dévoyer, a au contraire permis des confrontations d'hypothèses extrêmement fructueuses entre la psychologie et l'éthologie : on peut par exemple en voir la trace dans [Minsky 1988], qui s'attaque à la psychologie en partant d'hypothèses similaires à celles de Lorenz (sans jamais le citer). Mais l'inverse est aussi vrai : l'éthologie a bénéficié de l'apport théorique de la psychologie cognitive, notamment sur les mécanismes de mémorisation ou de perception, grâce à l'usage qu'elle a su faire de ce vocabulaire.



Modèle d'activation hydromécanique de Lorenz

(d'après [Lorenz 1984])

---

[13] Un exemple en est le "déclenchment à vide" d'un comportement, survenant en l'absence des déclencheurs habituels, fréquemment observé chez les animaux et l'homme.

[14] Acquises par renforcement, apprentissage ou conditionnement.

[15] Il suffit pour s'en convaincre d'écouter les propriétaires d'animaux domestiques parler du comportement de leur chat ou chien. Combien de fois un vocable intentionnel purement spéculatif quant au capacités réelles de l'animal en question est-il employé ?

[16] Voir [Drogoul 1993] pour plus de détails.

Cependant, le modèle de Lorenz[17], s'il parvient à expliquer la causalité de déclenchement d'un comportement, est inadéquat dès lors qu'il s'agit d'exprimer le contrôle sur l'ensemble des comportements qu'un animal peut effectuer, ce qui fait appel à des catégories intentionnelles plus vastes : comment exprimer la notion de but, comment interagissent des motivations contradictoires, etc. C'est à cette tâche que s'est attelé N. Tinbergen [Tinbergen 1971], qui a construit, en partant du modèle de Lorenz, son schéma hiérarchique de comportements. Ce schéma se caractérise par une classification des comportements (depuis les comportements moteurs, en bas de l'échelle, jusqu'aux fonctions - la reproduction, par exemple) en une hiérarchie fondée sur des liaisons d'activations verticales (les motivations, du haut vers le bas de l'échelle) et des liaisons d'inhibition horizontales. Les comportements enserrés dans cette hiérarchie sont bien sûr soumis à leurs déclencheurs externes et peuvent être partagés entre plusieurs hiérarchies (représentant plusieurs fonctions). L'intérêt de ce modèle, qui a été repris un grand nombre de fois en IA , est de proposer une vision d'ensemble de ce que peut être le comportement d'un agent soumis à une grande variété de stimulations externes et de besoins internes, sans préjuger de ses capacités cognitives. Que R. Brooks, pour son architecture de subsumption [Brooks 1990] (qui n'est qu'une variante simplifiée de ce schéma) et [Simon 1982] s'y référent tous deux pour décrire le comportement de leurs agents en dit long sur sa généralité - et sur l'originalité des "découvertes" en IA ou Vie Artificielle (voir [Maes 1991], qui fait le lien entre ce modèle et celui de M. Minsky).

Outre les modèles de comportement, beaucoup de travaux en éthologie peuvent susciter des interrogations intéressantes pour la conception d'agents artificiels, en proposant, toujours grâce à la démarche méthodologique que nous avons mentionnée, des choix d'implémentation alternatifs. Pour ne prendre qu'un exemple, considérons la notion de mémoire. Faut-il systématiquement la considérer comme une internalisation symbolique de représentations ou bien doit-on, dès lors que l'agent est plongé dans un environnement, également considérer la mémoire externe qu'il peut constituer par son activité ? L'éthologie invite à embrasser ces deux formes de maintien d'informations dans le temps sous un même concept ce qui permet, sans forcément aller jusqu'au quasi-béhaviorisme d'un R. Brooks, de faire l'économie, dans bien des cas, d'une représentation coûteuse de l'environnement et des traitements associés (voir par exemple [Drogoul 1993 - ch. 9 et 10][Beer 1990]).

3. Organisations

A partir de l'instant où la question de l'*organisation* d'agents artificiels est posée, il est légitime de rechercher et de confronter les modes d'organisation qui, dans la réalité, permettent à des agents naturels de fonctionner collectivement : de l'auto-organisation physico-chimique aux organisations rationnelles que nous connaissons dans nos sociétés, tout mode de composition devient la métaphore possible d'une procédure de fonctionnement collectif [Lestel et al. 1994].

---

[17] qui a, depuis, été largement revu et corrigé (voir par exemple [Pearce 1987]).

A cet effet, la sociologie animale (ou l'éthologie sociale) [Chauvin 1982; Tinbergen 1974] propose un certain nombre de pistes qui ne demandent qu'à être explorées. En étudiant la façon dont des agents naturels, qui n'ont pas forcément *conscience* de l'organisation à laquelle ils appartiennent, parviennent à coopérer, elle fournit, à l'instar de l'éthologie et de ses modèles de comportement, des modèles d'organisation souvent très simples, et incomparablement plus riches[18] que ceux auxquels nous ont habituées la sociologie ou la psycho-sociologie : symbiose, parasitisme, auto-organisation, hiérarchies, commensalisme, mutualisme, etc.

3.1 *Modèles intentionnels de fonctionnement collectif*

Ici encore, la démarche méthodologique réductionniste/intentionnaliste s'applique avec bonheur : elle permet de proposer des modèles de coopération, de coordination ou d'organisation, c'est-à-dire des modèles *intentionnels* de fonctionnement collectif, qui reposent sur des mécanismes d'interaction clairement identifiés et qui ne nécessitent souvent pas d'hypothèse de rationalité individuelle. L'intentionnalité ici est clairement attribuée, mais pas sur n'importe quelle base : comme toute catégorie intentionnelle doit pouvoir servir de trame dans un schéma explicatif fonctionnel, la question de sa définition et de sa fonction est posée d'emblée. Utiliser le même terme, par exemple *hiérarchie*, pour qualifier l'ensemble des relations inter-indiviuelles de dominance aussi bien chez les fourmis [Hölldobler et Wilson 1990] que chez les primates [De Waal 1982; Thierry 1990] ne relève pas de la vague généralisation, mais procède d'une définition rigoureuse de la fonction d'une hiérarchie pour les espèces sociales: réduction des conflits, centralisation de la décision, optimisation de la reproduction, etc. [McFarland 1990 - art. Hiérarchie][Armitage 1987]. Rapprocher ainsi ces deux occurrences d'organisation sociale permet d'appréhender les convergences et les différences quant aux mécanismes proximaux (agressions, inhibition, etc.) qui les produisent, et ainsi de mettre en lumière la diversité des moyens avec lesquels "Dame Nature" (pour reprendre le terme de D. Dennett) a pu concevoir des hiérarchies - comportements collectifs tendus vers un même but et assurant un succès reproductif qualitativement comparable. Les fourmis ont peut-être un comportement rigide, les primates sociaux font sans doute preuve d'une plus grande adaptabilité, mais, au bout du compte, quelle est la différence ? Du point de vue de la catégorie intentionnelle étiquetée "hiérarchie" et de la fonction qu'elle supporte, elle apparaît minime.

3.2 *Application à l'I.A.D.*

Même s'il s'agit de la même logique, une telle démarche en IAD serait bien entendu inverse : si le concepteur d'un système souhaite obtenir une organisation

---

[18] par exemple parcequ'ils peuvent associer plusieurs espèces distinctes, ce qui n'est pas le cas de la sociologie.

artificielle qui, observée attentivement lors de son fonctionnement, possède toutes les caractéristiques d'intentionnalité requises, et s'il dispose de plusieurs solutions technologiques possibles pour l'implémenter, le bon sens voudrait qu'il choisisse la plus simple (c'est-à-dire qu'il applique le choix de moindre complexité individuelle [Connah 1991; Parunak 1996]) ! Curieusement, c'est souvent le choix inverse qui est effectué - en témoigne la profusion d'architectures type BDI (beliefs-desires-intentions) ou de langages de communication complexes (comme KQML) pour implémenter, le plus souvent, des organisations d'une pauvreté criante[19].

La critique souvent faite à ce genre d'arguments considére qu'une organisation d'agents simples ou réactifs ne fait en réalité qu'imiter grossièrement les *vraies* organisations, obtenues avec de *vrais* agents (sous-entendu: des agents rationnels). Cela ressemble un peu à l'affirmation qui consiste à dire que les fourmis, par opposition aux hommes, ne peuvent pas coopérer au plein sens du terme parce qu'elles ne font que suivre un programme aveugle - ce qui mériterait déjà d'être démontré. Mais sur quels autres critères se basent les critiques pour affirmer de telles choses ? Si la *fonction* collective réalisée est qualitativement semblable, si elle permet au système de résoudre les mêmes problèmes d'une manière équivalente, pourquoi ne pas ranger ces comportements collectifs dans la même catégorie *intentionnelle* : seuls les mécanismes proximaux, qui de toutes façon découlent d'une intentionnalité originelle, celle du concepteur, différent! Cette intentionnalité originelle, quelque soit la manière de coder les agents, sera de toutes façons présente dans le système, sous une forme ou une autre. Pourquoi alors vouloir la masquer ? Comme nous l'avons montré plus haut, il n'y a aucune espèce de contradiction entre concevoir un système multi-agent *coopératif* et l'implémenter sous forme d'agents réactifs, car ce sont deux niveaux de conception distincts, de même que l'analyse causale et l'analyse fonctionnelle en éthologie sont deux niveaux d'analyse distincts.

L'important est de comprendre que le comportement collectif de coopération d'un ensemble d'agents en interaction ne sera **jamais** *intrinsèquement intentionnel*, mais qu'il rentrera (ou ne rentrera pas) dans la catégorie intentionnelle étiquetée "coopération" de l'observateur du système. Ainsi, contrairement à ce que pensent [Brassac et Pesty 1996], les "agents cognitifs" n'ont pas plus d'intentionnalité propre concernant la coopération que les agents réactifs (ou pas moins, selon le point de vue) : nous ne faisons que leur attribuer cette intentionnalité, et, par là même, nous ne faisons qu'interpréter l'intentionnalité originelle que le concepteur a bien voulu placer dans son système. Bien entendu, cette attribution ne s'avère justifiée que lorsqu'elle fonctionne, c'est-à-dire lorsqu'elle permet de comprendre, d'expliquer et de prévoir le comportement collectif des agents.

Si nous admettons donc que nous pouvons concevoir des organisations d'agents fonctionnelles et coopératives, que nous pouvons utiliser dans la démarche de

---

[19] un bon exemple en est la multitude de travaux de ce type autour de l'application proie-prédateurs ou du monde des cubes, alors que [Korf 1991] et [Ferber 1989] ont montré que des organisations d'agents simplissimes, non seulement résolvaient ces deux problème, mais étaient également capable de bluffer n'importe quel observateur - c'est-à-dire que celui-ci ne pouvait les distinguer qualitativement d'une organisation d'agents plus complexes.

conception des termes comme coopération, organisation sociale, coordination sans avoir à systématiquement passer sous les fourches caudines de la rationnalité individuelle, l'éthologie à quelques modèles intéressants à nous proposer.

Des modèles d'organisation sociale, tout d'abord. Les organisations les plus complexes se sont développées chez les insectes et les invertébrés, et les plus exploitées en IAD [Deneubourg & al. 1992; Drogoul 1993], sans doute en raison du grand nombre d'études dont elles ont fait l'objet en biologie [Wilson 1985], sont les sociétés d'hyménoptères: fourmis, abeilles, guêpes, bourdons, etc.[Hogeweg et Hesper 1982; Theraulaz 1991; Corbara 1991]. Il faut préciser que ce sont là des animaux sociaux par excellence, puisque l'ordre des hyménoptéres a connu onze développements indépendants de la socialité durant son évolution (plus qu'aucun autre ordre). On retrouve dans ces sociétés des types d'organisations avancés telles que les hiérarchies, la division du travail, des coordinations spatio-temporelles complexes, etc. Mais de nombreux autres animaux ont développé ces mêmes capacités, depuis les invertébrés coloniaux (coraux, syphonophores, bryozoaires, formés de "zooïdes" capables de se différencier fonctionnellement), jusqu'aux vertébrés, parmi lesquels des modèles fascinants peuvent être étudiés : les bancs de poissons ou les bandes d'oiseaux dans lesquels existent des degrés certains de coordination dans le comportement, que ce soit pour le mouvement ou les activités, ou bien encore les sociétés de primates, qui pourraient constituer une base idéale d'inspiration pour les systèmes ouverts, où une forme de cognition sociale artificielle semble nécessaire [Castelfranchi et Conte 1992]. La liste est loin d'être exhaustive, mais il nous a semblé utile de préciser que l'apport de l'éthologie sociale à la conception de systèmes collectifs artificiels ne doit pas se limiter qu'aux seuls insectes sociaux...

Des modèles de coopération, ensuite. Nous les distinguons des modèles d'organisation sociale, car une grande partie des études effectuées en éthologie sur la coopération porte sur des coopérations entre individus d'espèces différentes, c'est-à-dire sur les phénomènes de symbiose. Ces modèles sont intéressants à plus d'un titre pour l'IAD, dans le sens où ils proposent des voies originales pour marier et faire coopérer des agents dont l'architecture, les modes de communication et les modes d'interaction sont radicalement différents. Ironiquement, les relations symbiotiques semblent avoir évolués, non pas à partir d'attitudes altruistes, mais grâce au parasitisme, lui-même provenant de situations dans lesquelles des organismes envahissaient l'habitat d'autres animaux. Les modèles existants sont également diversifiés en ce qu'ils considèrent aussi bien les relations symbiotiques individus/individus, que des relations individus/sociétés, ou sociétés/sociétés. Ils constitueront dans les prochaines années, à n'en pas douter, un champ d'investigation particulièrement riches pour qui s'intéresse à faire coopérer des agents hétérogènes.

4. Conclusion

Le point essentiel de cet article est que nous devons prendre au sérieux, dans la conception de nos systèmes, qu'il n'y a pas d'attribution fonctionnelle ou de sens, explicite et objective, mais seulement des attributions de croyances dans le contexte d'une interprétation du fonctionnement du système. Ceci est compatible avec toute implémentation particulière; que les agents aient des états internes *intrinsèques* ou non, ils auront une intentionnalité constituée des désirs et des croyances que l'observateur leur attribuera. Et ces attributions trouveront leur légitimité dans le fait qu'elles fonctionneront, c'est-à-dire qu'elles lui permettront d'expliquer, de comprendre et de prédire le comportement de ces agents.

Le second point est que des descriptions fonctionnelles appliquées à des agents non-rationnels, ou à des processus aveugles, ne constituent en aucune manière une hérésie : correctement gérées, elles mettent en lumière des aspects intéressants de ce qui est encore considéré comme une caractéristique propre aux agents rationnels, à savoir l'intentionnalité, et permettent donc d'envisager, pour la réalisation d'une fonction (individuelle ou collective) donnée, des choix d'implémentation alternatifs, où cette intentionnalité ne sera pas forcément représentée en tant que pièce du système.

Reste bien entendu la difficulté à accepter ces points, qui ne sont après tout que des hypothèses de travail. Il ne fait aucun doute qu'ils le seront d'autant mieux que des méthodologies de conception orientée agent permettront de réellement les prendre en compte dans la construction d'un système. C'est en partie le sens de notre travail sur la méthode *Cassiopée*, détaillée dans [Collinot et al. 1996], méthode dans laquelle le point de vue du concepteur et celui de l'agent, ou, si l'on préfère, le point de vue sur la conception et celui sur l'implémentation sont clairement distingués : le concepteur peut manipuler des entités abstraites tels que coopération ou organisation dans l'analyse ou la description de son système, mais il est libre de les représenter ou non lors de son implémentation.

Bibliographie

ARMITAGE K. B., "Social dynamics of mammals : reproductive success, kinship and individual fitness", TREE, 2, 279-284, 1987.

BEER R. D., Intelligence as Adaptive Behavior, Academic Press, Londres, 1990.

BELEW R. K., "Artificial Life: A Constructive Lower Bound for Artificial Intelligence", IEEE Expert, February 1991, 1, pp. 8-15, 1991.

BRASSAC C. et PESTY, S. "La Pelouse Fourmilière : de la coaction à la coopération", in Actes de JFIADSMA'96, Hermes, Paris, 1996.

BROOKS R., "Elephants don't Play Chess", Journal of Robotics and Autonomous Systems, n°6, vol. 1,2, pp. 3-15, 1990.

BROOKS R., Intelligence without Reason, Proceedings of the IJCAI'91, Sydney (Australie), Morgan-Kaufmann, pp. 569-595, 1991.

CASTELFRANCHI C. ET CONTE R., Mind is not Enough: Precognitive Bases of Social Interaction, Proceedings of the Simulating Societies Symposium, University of Surrey, Guildford (Angleterre), pp. 93-110, 1992.

CHAUVIN R., Les Sociétés Animales, Presses Universitaires de France, Paris, 1982.

COLLINOT A., DROGOUL A., et BENHAMOU P., "Agent-Oriented Design of a Soccer Robot Team", in Proceedings of ICMAS'96, AAAI Press, 1996.

CONNAH D., "Why We Need a New Approach to the Design of Agents", AISBQ, 76, Spring 1991, pp. 13-15, 1991.

CORBARA B., *L'organisation Sociale et sa Genèse chez la Fourmi Ectatomma ruidum Roger*, Thèse de troisiéme cycle, Paris XIII - Villetaneuse, 1991.

DAVIES D. "Le béhaviorisme méthodologique et le réalisme intentionnel chez Dennett", in Lekton, D. Fisette (sous la dir. de), Vol. II, n°1, pp. 69-88, UQAM, 1992.

DAWKINS R., "Hierarchical Organization : a candidate principle for ethology", in *Growing Points in ethology*, Bateson & Hinde (eds), Cambridge Univ. Press, 1976.

DENNETT, D. , "La stratégie de l'interprète: le sens commun et l'univers quotidien", Paris, Gallimard, 1990.

DENNETT, D. "Darwin's Dangerous Idea", Touchstone, Simon & Schuster, New-York, 1995.

DE WAAL, F.B. Chimpanzee Politics: power & sex among apes. Harper & Row. 1982.

DROGOUL A., "De la Simulation Multi-Agents à la Résolution Collective de Problèmes: une Etude de l'Emergence de Structures d'Organisation dans les Systèmes Multi-Agents", Thèse de Doctorat de l'Université Paris VI, Novembre 1993.

DUMOUCHEL P. "Dennett et le béhaviorisme méthodologique", in Lekton, D. Fisette (sous la dir. de), Vol. II, n°1, pp. 53-68, UQAM, 1992.

ENGEL, P., "Introduction à la philosophie de l'esprit", La Découverte, Paris, 1994.

GUILLOT, A. Revue générale des méthodes d'étude des séquences comportementales, in Etudes et analyses comprtementales, vol. 2, n°3, pp. 86-106, 1986.

GRIFFIN, D.R. Animal mind - Human mind. Springer Verlag, 1982.

HOGEWEG P. ET HESPER B., "The Ontogeny of the Interaction Structure in Bumble Bee Colonies: a MIRROR Model", Behavioral Ecology and Sociobiology, 12, 271-283, 1983.

HÖLLDOBLER B. ET WILSON E. O., The Ants, Springer-Verlag, Berlin, 1990.

JAISSON P., *La Fourmi et le Sociobiologiste*, Editions Odile Jacob, Paris, 1993.

KORF R. E., *A Simple Solution to Pursuit Games*, Actes du Distributed Artificial Intelligence Workshop, Boston, pp. 183-194, 1991.

LESTEL D., GRISON B. ET DROGOUL A. "Les Agents Réactifs et le Vivant dans une Perspective d'Evolution Coopérative", in Intellectica, n°19, pp. 73-90, 1994.

LOEB J., *Comparative Physiology of the Brain and Comparative Psychology*, Putnam, New York, 1900.

LORENZ K., *Les Fondements de l'Ethologie*, Flammarion, Paris, 1984.

McFARLAND, D.J. Decision making in animals, in Nature, n°261, pp. 15-21, 1977.

McFARLAND, D.J., Animal Behaviour. Psychobiology, ethology & evolution. Longman Scientific & Technical, 1985.

McFARLAND, D.J., Dictionnaire du comportement animal, Robert Laffont, Paris, 1990.

MAES P., "A Bottom-Up Mechanism for Behavior Selection in an Artificial Creature", in From Animals to Animats, MIT Press, Cambridge, pp. 239-246, 1991.

MEYER J.-A. ET GUILLOT A., "From Animals to Animats: Everything you Wanted to Know about the Simulation of Adaptive Behavior", Rapport Technique du Groupe de BioInformatique de l'Ecole Normale Supérieure, n°90-1, 1990.

MINSKY M., *La Société de l'Esprit*, InterEditions, Paris, 1988
MORGAN C. L., *An Introduction to Comparative Psychology*, Scott, Londres, 1894.
PARUNAK, V. D., "Go to the ant", Proceedings of the ICMAS'96 Workshop on Animal Societies as an Alternative Metaphorical Basis for DAI.
PEARCE, J.M., An introduction to animal cognition. Lawrence Erlbaum Assoc., 1987.
RICHARD, G. Histoire des idées en Ethologie (1ère Partie), in Etudes et analyses comportementales, vol. 2, n°3, pp. 131-190, 1984.
RICHARD, G. Histoire des idées en Ethologie (2ème Partie), in Etudes et analyses comportementales, vol. 2, n°4, pp. 195-254, 1985.
SIMON, H.A., The sciences of the artificial, MIT Press, 1982.
SKINNER B. F., *The Behavior of Organisms*, Appleton-Century-Crofts, New York, 1938.
SOBER E., Methodological behaviorism, Evolution and Game Theory, in *Sociobiology and Epistemology*, J. Fetzer (Ed.), Dordrecht, Reidel, pp. 181-200, 1985.
THERAULAZ G., Morphogenèse et Auto-Organisation des Comportements dans les Colonies de Guêpes Polistes Dominilus (Christ), Thèse de Troisiéme Cycle, Université d'Aix-Marseille I, 1991.
THIERRY B., "Feedback Loop between Kinship and Dominance: the Macaque Model", *J. Theor. Biol.*, vol. 145, pp. 511-521, 1990.
TINBERGEN N., *L'Etude de l'Instinct*, Editions Payot, Paris, 1971.
TINBERGEN N., La Vie Sociale des Animaux, Payot, Paris, 1979.
VAUCLAIR J., *L'intelligence de l'Animal*, Editions du Seuil, Paris, 1992.
WATSON J. B., *Psychology from the Standpoint of a Behaviorist*, Lippincott, Philadelphie, 1919.
WILSON E.O., "The Sociogenesis of Insect Societies", Science, n°228, pp. 1489-1495, 1985.