

Un animat coopératif pour la simulation des comportements adaptatifs de l'animal

Samuel Delepouille^{*,}**

^{*} URECA

UPRES-EA 1059

Université de Lille 3

B.P. 149

59653 Villeneuve d'Ascq Cedex

France

delepouille@univ-lille3.fr,

darcheville@univ-lille3.fr

Philippe Preux^{}**

Jean-Claude Darcheville^{*}

^{**} LIL

Université

du Littoral Côte d'Opale

B.P. 719

62228 Calais Cedex

France

preux@lil.univ-littoral.fr

Résumé

Dans cet article, nous décrivons MAABAC, un système qui repose sur la sélection des comportements par leurs conséquences. MAABAC simule la dynamique d'un bras basé sur l'activité de ses muscles. L'architecture de MAABAC est totalement décentralisée; chaque muscle reçoit peu de stimulus de l'environnement et décide de son action sur cette base.

Nous portons un grand intérêt à la comparaison entre la dynamique démontrée par MAABAC et le développement des mouvements d'atteinte chez le jeune enfant. Malgré sa simplicité, MAABAC démontre des comportements tels que la généralisation et l'extinction; il est également sensible au shaping, une technique bien connue de l'analyse du comportement des organismes.

Une conséquence de notre travail est qu'il donne des arguments à l'étude du comportement basée sur des principes sélectionnistes. MAABAC apprend des mouvements d'atteinte sans utiliser aucun système cognitif. Un autre point clé de ce travail est qu'il rend compte du développement du comportement qui commence par des mouvements non organisés.

1. Introduction

L'un des apports majeurs de la perspective animat a été de montrer que des comportements complexes tels que ceux des organismes vivants peuvent résulter de l'interaction d'éléments relativement simples. C'est d'ailleurs cette relative simplicité des éléments qui le composent qui nous permet de faire une implantation réelle des animats. Les mécanismes d'émergence sont particulièrement importants pour cette science jeune dont le but est de mieux comprendre les mécanismes de la vie telle qu'elle

est en la comparant à des animaux artificiels, volontairement simplifiés mais présentant les structures essentielles à l'adaptation.

En implantant certains processus, nous pouvons contribuer à réduire la complexité de phénomènes naturels en montrant qu'elle peut résulter de la combinaison de processus plus simples. Ainsi, l'implantation d'algorithmes adaptatifs basés sur des principes de renforcement peut suggérer que des phénomènes *a priori* complexes se laissent réduire à des processus de sélection en interaction, soit des processus de co-évolution, mettant en avant l'importance des processus de sélection et de l'interaction d'un agent avec son environnement. Que ce soit par Darwin et ses successeurs en biologie, Skinner en analyse du comportement, ou encore Edelman en neurophysiologie, les processus sélectionnistes sont invoqués pour expliquer de nombreux processus naturels importants; en les implantant avec précaution, il est parfois possible de montrer qu'une certaine observation peut résulter d'un (simple) processus sélectionniste ou de la combinaison de processus sélectionnistes (voir (Mitchell and Belew, 1996)). Ainsi, les travaux sur les « univers virtuels » montrent que des structures complexes peuvent évoluer à partir de règles simples (voir notamment les travaux de T. Ray sur *Tierra* (Ray, 1994), ou les travaux autour de *Echo* (Forrest and Jones, 1994, Cuvelier et al., 1997)). Par ailleurs, on a fait évoluer des réseaux de neurones capables de contrôler des animats voire, exhiber des structures modulaires (Calabretta et al., 1998). D'autres travaux montrent que des phénomènes complexes du domaine du vivant peuvent s'expliquer à l'aide d'algorithmes de renforcement (Spier and McFarland, 1998, Delepouille et al., 2000) et des interactions avec le milieu (Scheier and Pfeifer, 1998, Touretzky and Saksida, 1996).

Le but de cet article est de montrer que des résultats concernant l'organisation de comportements collectifs peuvent se montrer intéressants pour la création d'animat. La notion d'émergence (Steels, 1994), bien qu'omniprésente dans la littérature animat, ne doit pas nous amener à ignorer la structure des interactions qui lient les composants d'un système pour aboutir au comportement global que nous observons. C'est cette évolution de la structure des comportements coopératifs que nous nous proposons d'étudier au travers de la notion de contingence coopérative. Ainsi, notre travail s'inscrit dans la perspective de l'appariement entre système naturels et artificiels puisque nous tentons de créer des systèmes artificiels dont les comportements sont sensibles à la dimension historique comme on l'observe dans les systèmes complexes naturels. Nous montrerons donc dans cet article que des observations menées sur l'évolution de dyades d'animaux peuvent nous renseigner sur l'évolution de systèmes coopératifs; nous décrirons ensuite l'architecture d'un système coopératif MAABAC (Multi Agents Animat for Behavioral Arm Control) dans un environnement virtuel. Nous montrerons enfin que le système MAABAC présente une évolution comparable à celle de systèmes naturels tels qu'ils ont été décrits par les spécialistes du comportement. Les comportements du système MAABAC sont comparables aux propriétés d'apprentissage observés chez l'animal ou chez le bébé humain (évolution et robustesse de l'apprentissage, extinction, sensibilité au contexte, modelage du comportement). Ces propriétés sont produites par une architecture entièrement parallèle et émergent de cette organisation sans avoir été explicitement programmées pour obtenir ce résultat.

2. Observation et résultats de système coopératifs naturels

Dans cette section, nous montrerons que des observations réalisées sur l'organisation de systèmes coopératifs naturels peuvent nous guider pour réaliser des systèmes artificiels. Dans un premier temps, nous montrons qu'une analyse basée sur la sélection du comportement permet une analyse des phénomènes coopératifs entre individus. Dans un deuxième temps, nous utilisons les résultats obtenus pour modéliser un système complexe naturel composé d'agents coopératifs.

2.1 coopération entre individus, une approche basée sur la sélection

La question de l'évolution de la coopération a suscité de nombreux débats dans l'ensemble de la communauté scientifique. Cette question a d'abord été envisagée en référence à la théorie des jeux (Axelrod, 1984) dans le cadre bien connu du Dilemme Itéré des Prisonniers.

D'autres auteurs se sont cependant interrogés sur l'intérêt écologique d'une telle description (Hemelrijk, 1997, Killingback et al., 1999).

Les problèmes liés à la dynamique du comportement nous paraissent essentiels. Nous avons proposé (Delepoulle et al., 2000) d'utiliser une situation basée sur la modélisation des comportements coopératifs. Cette situation est connue dans la littérature sous le nom de situation sociale minimale (Sidowski et al., 1956, Sidowski, 1957).

Avant d'aller plus loin, il est essentiel de définir ce que nous entendons par coopération en terme de comportements. La notion comportementale de coopération fait référence aux contingences de renforcement qui traduisent la conséquence pour l'organisme d'un comportement dans un état donné de l'environnement. La contingence coopérative est caractérisée par le fait que le renforcement *du comportement* d'un sujet dépend au moins en partie du comportement d'un autre sujet (Hake and Vukelich, 1972). Dans ce cadre, un comportement coopératif n'a pas forcément de conséquence positive pour l'organisme qui en est à l'origine. Cependant, ce comportement coopératif peut être à l'origine d'une conséquence positive pour un autre organisme. C'est précisément pour cette raison que les comportements coopératifs ont paru si intéressants à étudier. Ils paraissent en quelque sorte en contradiction avec une approche sélectionniste puisqu'ils se maintiennent en l'absence de conséquences directement positives.

De nombreuses recherches (Hake and Vukelich, 1972, Hake and Vukelich, 1973, Schmitt, 1976, Hake and Olvera, 1978) ont tentées de comprendre comment les comportements coopératifs pouvaient apparaître et se maintenir. Ces auteurs insistent sur l'importance de l'interdépendance des réponses qui garantit leur maintien dans le temps. Dans ce cas, la réponse coopérative d'un sujet est sous la dépendance de la réponse coopérative de l'autre sujet et réciproquement. Hake indique que cette situation, qu'il désigne sous le terme de **coopération maximale** se traduit par une augmentation de la correspondance entre les réponses coopératives. Pour cet auteur, la présence d'une contingence coopérative est marquée par le développement d'une coordination entre les réponses des organismes placés dans cette contingence. Si nous sommes bien en présence d'une situation de coopération maximale, la diminution des réponses coopératives de l'un des agents devrait provoquer également une diminution des réponses de l'autre. C'est pourquoi Hake suggère de mesurer la coopération en rapportant le nombre de réponses coopératives non pas à des unités de temps mais au nombre de réponses de l'autre agent.

Ces résultats vont dans le même sens que ceux que nous obtenons en situation sociale minimale. Nous avons montré (Delepoulle et al., 2000) qu'il existe une interac-

tion positive entre la contingence coopérative et la synchronisation des comportements. L'apparition à un moment donné de la coopération renforce la synchronisation et réciproquement, la synchronisation inter sujets est facilitée par la coopération.

2.2 Utilisation de ces résultats pour des systèmes coopératifs artificiels

L'ensemble de ces résultats ainsi que nombre autres travaux (Maes and Brooks, 1990, Mataric, 1995, Mataric et al., 1998) laissent penser qu'il est possible de produire des comportements coordonnés et adaptatifs en prenant pour base des agents relativement simples et autonomes les uns des autres, sans forcément recourir à une planification centrale.

La synchronisation des comportements d'agents par le biais de l'organisation de contingences coopérative doit permettre de réaliser de nombreux systèmes autonomes. Nous avons choisis, à titre d'exemple de réaliser le système MAABAC qui modélise le comportement d'un bras mobile. L'intérêt est de pouvoir comparer l'évolution de comportements tels que l'atteinte ou la poursuite en les comparant aux résultats de bébés humains qui réalisent leurs premières atteintes.

3. L'architecture MAABAC, un animat composé de systèmes coopératifs pour le contrôle de bras mobile

MAABAC consiste en un bras contrôlé par des muscles, et un système visuel.

Dans un premier temps, notre propos n'est pas de modéliser fidèlement l'architecture d'un bras. En effet, le bras humain comporte plusieurs centaines de muscles qui interagissent entre eux. Nous utilisons ici une simplification compatible avec les travaux des neurosciences qui montrent que les muscles agissent au sein de groupes cy-négétiques (Rosenzweig and Leiman, 1989). Les simplifications que nous faisons sont acceptables car nous voulons essentiellement démontrer qu'une architecture non centralisée peut présenter des comportements coordonnés relativement proches de ceux observés réellement. Dans ce sens, nous recherchons plus une similarité dans le comportement que dans la structure qui est à l'origine de ces comportements.

La modélisation que nous proposons envisage la commande musculaire d'un bras, comme un système multi-agents. Chaque agent possède essentiellement une perception locale du muscle sur lequel il intervient. Dans cette section, nous allons présenter le fonctionnement général de la plateforme de simulation, l'architecture MAABAC puis le fonctionnement des agents du système.

3.1 La plateforme de simulation pour MAABAC

La plate-forme de simulation, développée en Javatm est destinée à modéliser le comportement d'un bras simplifié en deux dimensions. Cette plate-forme permet d'implémenter l'architecture MAABAC (Multi-Agent Animat for Behavioral Arm Control). La représentation en deux dimensions est ici suffisante dans la mesure où nous voulons savoir si l'architecture MAABAC peut aboutir à des comportements organisés. La plate-forme de simulation permet à l'architecture MAABAC de contrôler le bras constitué de deux segments caractérisés par leur longueur (l_1 et l_2) et la valeur angulaire des deux articulations α et β . La figure 1 indique la position du bras en fonction de ces différents paramètres.

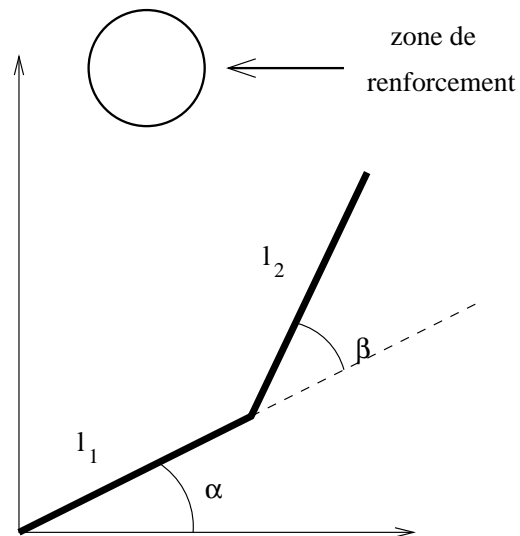


FIG. 1 – Modélisation de la position d'un bras en deux dimensions. Le bras est composé de deux segments de longueur respectivement l_1 et l_2 , ces deux segments étant joints par une articulation (un « coude »). L'extrémité du bras qui correspond à l'épaule est immobile. La position du bras est déterminée par deux angles, α entre l'horizontale et l'avant-bras, β l'angle entre les deux segments du bras (le « coude »). Dans l'expérience décrite ici du développement d'un mouvement coordonné d'atteinte, l'extrémité libre du bras doit atteindre et demeurer dans la zone de renforcement circulaire. Celle-ci peut changer de taille et de position.

D'autre part, le simulateur permet également de fixer les conditions initiales, l'intensité du renforcement, l'étendue de la zone de renforcement et le coût des mouvements. Il est également possible d'enregistrer les mouvements produits par l'animat.

3.2 L'architecture MAABAC

Dans le système MAABAC, chaque muscle est contrôlé par un agent, comme le montre la figure 2. Cet agent correspond à l'entité biologique qui effectue la commande musculaire (motoneurones, structure centrales, ...). Chaque agent contrôle un et un seul muscle. Les muscles sont au nombre de deux par articulation. L'un provoquera le fléchissement de l'articulation, ce qu'on désigne habituellement sous le nom d'agoniste. L'autre muscle, désigné sous le nom de muscle antagoniste est responsable de l'extension de la même articulation. Ainsi la position d'une articulation est définie par la combinaison de l'activation de deux muscles agissant en sens contraire.

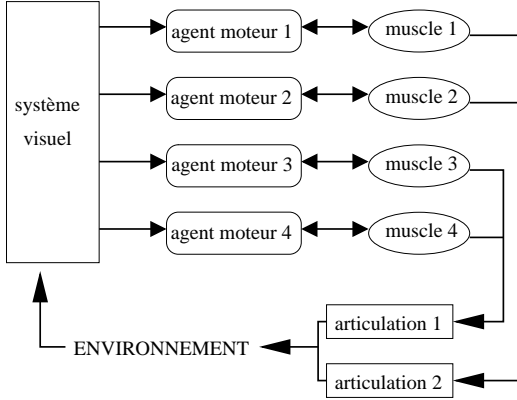


FIG. 2 – L'architecture MAABAC (voir le texte).

L'ensemble de ces comportements provoquent des modifications dans l'environnement. Le système visuel est sensible à ces modifications et donne à chaque agent la même information. Ce message spécifie de façon extrêmement simplifiée la distance et la position de la zone de renforcement relativement à l'extrémité du bras.

3.3 Les agents qui composent MAABAC

Nous présentons ici l'architecture des agents de l'architecture MAABAC, leur capacités perceptives, leurs comportements ainsi que l'algorithme d'apprentissage retenu. Cette organisation est représentée par la figure 3.

3.3.1 Perception

D'un point de vue perceptif, l'agent musculaire perçoit l'état de contraction du muscle et il reçoit également des afférences du système visuel, sous la forme de deux stimulations spécifiant la direction et la distance relative entre la main et la zone de renforcement. En fait, l'information visuelle transmise à l'agent de commande musculaire est relativement pauvre : la stimulation spécifiant la direction ne peut prendre que quatre valeurs

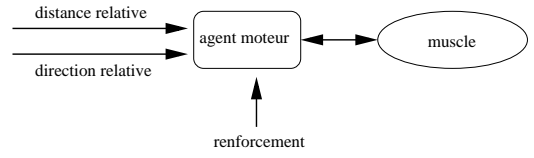


FIG. 3 – Perception et comportement d'un agent musculaire. Un muscle peut être relâché, contracté ou maintenir la tension courante. Dans chaque cas, l'intensité du comportement varie de 0 à 49. Cette intensité détermine le coût (énergétique) du mouvement. Le système visuel de MAABAC fournit deux stimuli, chacun pouvant prendre 4 valeurs (voir la figure 4).

qui correspondent aux quatre quadrants possibles pour la position angulaire de la zone de renforcement par rapport à la main. De même, la stimulation dépendant de la distance ne comprend que quatre valeurs. La figure 4 montre comment la position relative du bras et de la zone de renforcement est perçue. Chaque agent musculaire peut accéder à cette information mais elle ne détermine pas à elle seule le comportement de l'agent. En fonction de cette information de position et de l'état du muscle, l'agent prend une décision autonome en terme de contraction ou de relâchement.

3.3.2 Action

Les agents fonctionnant en parallèle, le comportement d'un agent n'affecte l'état que d'un seul muscle. Chacun des agents peut modifier l'état du muscle par le biais de trois comportements (contracter, relâcher ou maintenir la tension en cours). Un muscle est alors caractérisé par cinquante états de contraction. L'état zéro correspond au relâchement total du muscle alors que le cinquantième état représente une contraction maximale. En fonction de cet état, on définit également une fonction de consommation qui associe à chaque état une valeur modélisant la dépense énergétique pour maintenir cette contraction. Dans notre cas, nous avons représenté la dépense sous forme par la fonction quadratique de l'état :

$$\text{coût} = \kappa \left(\frac{\epsilon}{N} \right)^2 \quad (1)$$

où ϵ est l'état courant ($0 \leq \epsilon \leq N$).

De cette manière, les faibles contractions correspondent à des dépenses énergétique très modestes alors que les fortes contractions engendrent une consommation très importante.

3.3.3 Apprentissage

Nous voulons que les agents moteurs aient une organisation compatible avec les propriétés de plasticité du système central tel que l'a montré Edelman. Pour cela, les agents implantent l'algorithme d'apprentissage des

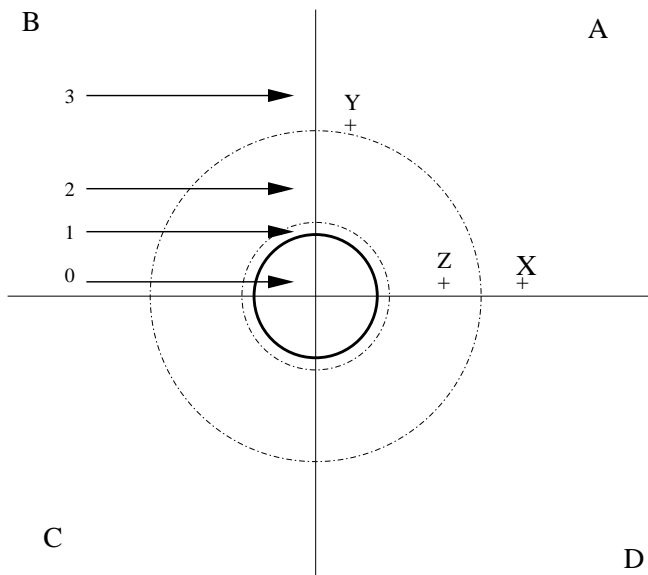


FIG. 4 – Perception de l'extrémité du bras en fonction de la zone de renforcement. L'espace est décomposé en 4 zones qui déterminent la distance perçue par le bras : 0 pour dans la zone de renforcement, 1 correspond à une position très proche, 2 à un éloignement modéré, 3 au reste de l'espace. Ainsi, les distances 0 et 1 correspondent à la région fovéale de la rétine, les zones 2 et 3 aux régions périfovéales. De même, la direction relative entre l'extrémité du bras et la zone de renforcement est décomposée en 4 quadrants (A, B, C et D). Par exemple, les positions X et Y sont perçues comme identiques, alors que les positions X et Z sont perçues comme différentes.

valeurs Q . En utilisant cet algorithme d'apprentissage, l'agent tente d'optimiser la quantité de renforcement reçue sur une longue période de temps. Comme le prévoit l'algorithme d'apprentissage des valeurs Q , le renforcement à long terme est pris en compte mais subit une dévaluation. La quantité de renforcement r est donnée par :

$$r = \text{reinf} - \text{cost} \quad (2)$$

où reinf vaut 1 si l'extrémité du bras est située dans la zone de renforcement, 0 sinon, et le cost est donné par l'équation 1.

Plus précisément, cet algorithme a été proposé par Watkins (Watkins, 1989) et est basé sur le Time derivative (TD) model de Sutton et Barto (Sutton and Barto, 1990). Itératif, cet algorithme consiste, en fonction de l'état courant et de sa valeur renforçatrice, à déterminer l'état suivant et à modifier la table des valeurs Q qui permettent à l'algorithme de prévoir le renforcement qu'il est susceptible d'obtenir dans un état donné, en émettant un certain comportement. L'émission du comportement c dans l'état e entraîne un renforcement r et le passage dans l'état e' ,

de quoi est déduite la modification des valeurs Q comme suit :

$$Q_{e,c} \leftarrow Q_{e,c} + \alpha(r + \gamma \max_d Q_{e',d}) - Q_{e,c} \quad (3)$$

α et γ sont des valeurs comprises entre 0 et 1.

Dans MAABAC, la table des valeurs Q est initialisée à 0. L'état courant est défini par l'intensité du comportement courant et les stimuli perceptifs. Le comportement à émettre est choisi dans 90% des cas comme celui qui maximise la prédiction du renforcement, au hasard sinon ; on a également utilisé une sélection du comportement proportionnelle à sa valeur Q . Ces deux techniques permettent une exploration de l'espace des comportements.

La position du bras dépend donc de l'état de contraction de chacun des agents musculaires. En fait, la position d'une articulation résulte de la combinaison linéaire des tensions musculaires des deux muscles de l'articulation (l'agoniste et l'antagoniste). D'un autre côté, le renforcement reçu par un agent dépend de l'ensemble des comportements des agents. Nous sommes donc bien dans une situation de coopération telle que définie dans un sens comportemental puisque le renforcement d'un agent dépend au moins en partie — en grande partie ici — du comportement de plusieurs autres agents.

Les résultats que nous avons observés quant à l'évolution de la coopération entre deux sujets humains dans une procédure minimaliste (Delepoulle et al., 2000) nous permettent de penser qu'un comportement organisé peut résulter d'une contingence de coopération. Nos résultats en situation minimaliste montrent en effet que la contingence coopérative aboutit à une situation de coordination. Il semble donc possible que la co-dépendance des agents musculaires pour le renforcement global soit suffisante pour induire une coordination et ce même en l'absence de coordination centrale. Telle est l'hypothèse que nous formulons.

4. Résultat du système MAABAC et comparaison avec le comportement de systèmes naturels

Le protocole expérimental est le celui : le bras est initialisé dans une position aléatoire ; la zone de renforcement est positionnée dans l'espace. On laisse l'algorithme évoluer jusqu'à ce qu'il atteigne la zone de renforcement. On laisse alors l'algorithme « apprendre » la position de la zone de renforcement pendant un certain nombre d'itérations. Lorsqu'il est demeuré un certain temps dans cette zone, on ré-initialise la position du bras.

4.1 Evolution du comportement

Au départ, le système présente un comportement d'apparence désordonné, comme le montre la figure 5. Ce

pattern correspond aux mouvements « spontanés » enregistrés chez le bébé au début de l'expérience. On observe ensuite que le bébé, au cours de ces mouvements spontanés, va faire passer sa main par hasard dans la zone de renforcement. C'est le cas pour le système MAABAC sur la figure 6. Le renforcement va produire une modification du comportement car ceux qui ont précédé l'arrivée du renforcement vont se produire plus fréquemment. La figure 7 montre qu'un comportement « volontaire » se met ensuite progressivement en place. En fait, les agents du système ne font que reproduire les comportements qui ont produit pour eux le renforcement, en fonction de la situation. Ce processus aboutit à un mouvement coordonné tel qu'on peut le voir sur la figure 8.

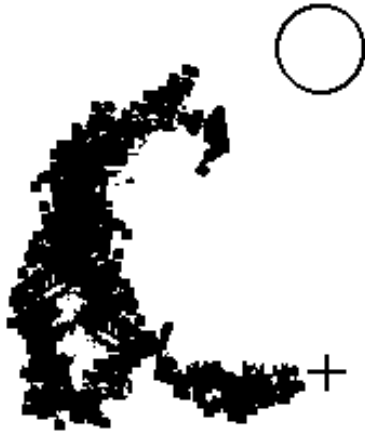


FIG. 5 – Représentation graphique du mouvement. Les différentes positions successives de l'extrémité du bras sont présentées sur la même figure. Au départ, le mouvement n'a pas d'organisation particulière et aboutit essentiellement à une exploration aléatoire de l'environnement.



FIG. 6 – Atteinte accidentelle de la zone de renforcement. Y étant renforcé, le bras a tendance à y demeurer.

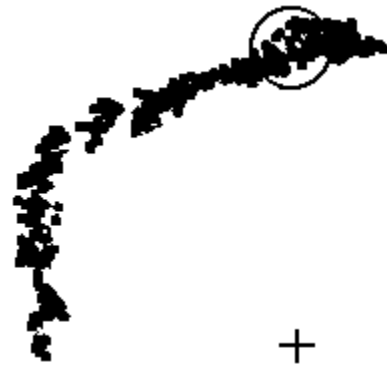


FIG. 7 – Première atteinte « volontaire ». Après avoir remis le bras en position initiale éloignée de la zone de renforcement, le bras retrouve plus rapidement la zone.

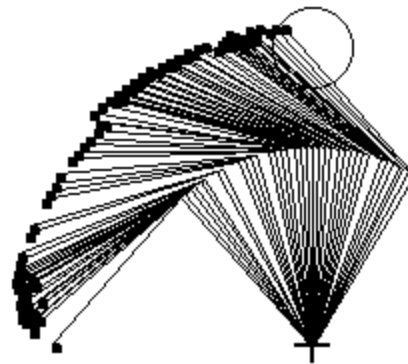


FIG. 8 – Mouvement coordonné. Après un certain nombre d'atteintes volontaires, un mouvement coordonné du bras est obtenu. On a ici représenté les deux segments du bras et son extrémité. L'atteinte de la zone de renforcement est ici directe.

4.2 Les co-activations musculaires

L'apparition de mouvements coordonnés résulte de l'action synchrone des agents musculaires. L'émergence de cette coordination a été rendue possible, comme nous le suggérons, par l'existence d'une contingence coopérative au sein du groupe d'agents musculaires. Plus précisément, on observe qu'un mouvement résulte d'une coordination entre les muscles des deux articulations mais aussi par l'apparition d'une action alternée des muscles au sein de la même articulation. Cette action alternée entre le muscle agoniste et le muscle antagoniste résulte du coût que nous avons donné à la contraction musculaire. En l'absence de coût, on observe que chaque état de l'articulation peut être obtenu par de nombreuses configurations musculaires. Par exemple, si l'articulation est en position médiane, ceci peut résulter d'un relâchement des

deux muscles antagonistes ou par une contraction identique dans les deux muscles. Si l'on ajoute la fonction de coût, les agents doivent optimiser le renforcement global tout en minimisant leur propre déplacement. Ceci a pour conséquence que chaque état de l'articulation correspond à un et un seul état pour les couples de muscles antagonistes, d'où l'apparition d'actions alternées entre les muscles d'une même articulation. Comme cet apprentissage se fait en fonction de l'expérience de chaque agent, on observe que le système évolue progressivement. L'atteinte se produit d'abord par une co-contraction de l'ensemble des muscles. Une fois cette atteinte installée, il y a progressivement réduction de la dépense musculaire nécessaire à la même atteinte. Ce phénomène a également été observé chez l'humain. Par exemple, dans le cas du développement de la marche, il a été observé chez le jeune enfant "a high degree of co-activation of antagonistic muscles" (Forssberg, 1986). Ce n'est que progressivement que l'organisme apprend à effectuer le même mouvement en minimisant la dépense, c'est-à-dire par alternance d'activation des muscles antagonistes.

L'intérêt de ce résultat est de montrer qu'une telle alternance, classiquement attribuée au développement d'un système d'oscillateur (sorte d'horloge interne) peut très bien se développer en l'absence totale de coordination centrale. Le paragraphe suivant montre que le système présente également une organisation globale de l'apprentissage.

4.3 Robustesse de l'apprentissage et extinction

Il nous a semblé intéressant de comparer les propriétés de notre système avec celle qui ont été classiquement observées chez l'animal (Skinner, 1938). Pour cela, nous avons utilisé le programme d'apprentissage suivant. Dans la première condition A, le renforcement est donné pour une position précise de l'extrémité du bras. Après un certain nombre d'atteintes de la zone de renforcement, on modifie sa position pour la condition B. Ensuite, dans la condition C, on procède à une extinction puisque le renforcement n'est plus distribué, quel que soit le comportement. Enfin, on réintroduit la condition B.

La figure 9 montre le nombre d'itérations nécessaires à MAABAC pour atteindre la zone de renforcement depuis une position initiale identique pour tous les essais. On observe un apprentissage pour la condition A car progressivement le nombre de pas nécessaire diminue jusqu'à obtenir un mouvement coordonné.

Ensuite, la modification (condition B) de la zone de renforcement provoque d'abord un allongement du temps nécessaire pour atteindre la zone de renforcement. Il est d'ailleurs à noter que les premières performances sous la condition B sont moins bonnes que sous la condition A. Ce phénomène est également observé chez l'animal car un apprentissage préalable peut avoir un effet perturbateur sur un deuxième apprentissage. On observe

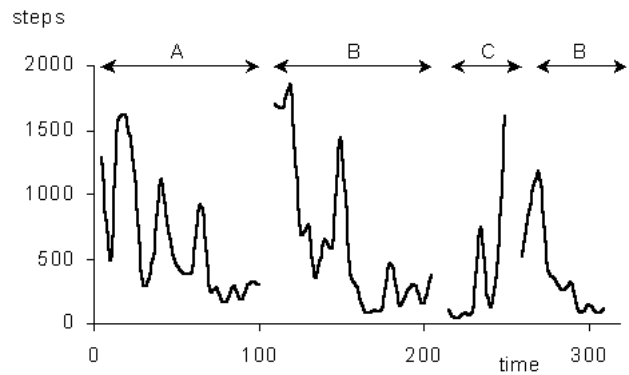


FIG. 9 – Évolution des performances de l'architecture MAABAC soumise à plusieurs programmes de renforcement. L'abscisse indique le temps, les ordonnées le nombre de d'itération de l'algorithme pour atteindre la zone de renforcement. Le programme A consiste en la présentation de la cible et renforcement lorsque l'extrémité du bras aboutit dans la zone de renforcement. Le programme B est identique suite à un déplacement de la zone de renforcement. Le programme C consiste à ne jamais renforcer le comportement (extinction). Enfin, le programme B est à nouveau introduit. L'économie lors du second programme B montre que le système est sensible à l'historique de l'environnement.

néanmoins que le critère d'apprentissage est atteint plus rapidement lors de la condition B. Ceci laisse penser qu'un phénomène de généralisation tel qu'il est observé chez l'animal est à l'origine de cette amélioration de la performance lors du second apprentissage.

Lors de la phase d'extinction (condition C), on observe que le comportement se maintient pendant un certain temps mais que progressivement il disparaît. L'intérêt de réintroduire ensuite la condition B est de montrer que l'extinction n'est pas une disparition de l'apprentissage car le comportement peut réapparaître très facilement. On observe bien lors de la ré-introduction du renforcement, identique à la première condition B, un phénomène connu sous le nom « d'économie au ré-apprentissage » observé dans l'apprentissage animal ??.

4.4 Modelage

Le modelage est une technique utilisée pour faire apprendre à un animal des comportements qui ont une très faible probabilité d'apparition spontanée. Pour cela, on commence par renforcer des comportements se rapprochant du comportement souhaité puis on augmente le degré d'exigence progressivement. Nous avons utilisé cette technique pour faire évoluer le comportement du bras. Pour cela, nous choisissons une zone de l'environnement qui n'est presque jamais explorée. C'est le cas de certaines zones extrêmes qui correspondent à une activation

maximale des muscles des deux articulations et le relâchement simultané des muscles antagonistes. Nous avons pu vérifier que cette zone n'est explorée en moyenne qu'une fois toutes les 20 000 itérations à partir d'un point de départ donné. Une tentative d'apprentissage s'en est révélée impossible car la consommation énergétique totale est trop élevée.

Un programme basé sur le modelage nous a cependant permis de faire apparaître un comportement d'atteinte directe à partir du même point de départ. Le programme consiste à positionner l'articulation à proximité de la zone de renforcement de telle sorte que la probabilité d'atteinte soit forte puis progressivement, on éloigne l'extrémité du bras de la zone de renforcement. Le programme que nous avons utilisé était constitué de 300 essais soit un total d'environ 11 000 itérations. À la fin de ce programme, la zone de renforcement était atteinte en une quarantaine d'itérations.

4.5 Comparaison avec des données expérimentales

Le système MAABAC supporte l'idée qu'un comportement comme le positionnement de l'extrémité d'un dispositif présentant de fortes analogies avec un bras, peut se stabiliser sur une position pour peu que cette dernière puisse avoir une valeur renforçatrice indépendamment des mouvements initiaux du bras. L'évolution du comportement de ce système n'est pas sans rappeler ce qui se passe au cours du développement du mouvement d'atteinte des objets (reaching) chez le jeune enfant. L'atteinte d'un objet est une question de positionnement correcte de la main dans l'espace. Rappelons que l'installation de ce mouvement s'étale, le plus souvent, entre le 4^e et le 8^e mois d'âge. Les travaux de Thelen (Thelen and Smith, 1994) font apparaître que le positionnement précis et exclusif de la main sur l'objet traduit la présence d'un attracteur point pour les patrons de mouvements de l'ensemble du bras qui initialement sont désordonnés par rapport à la cible. Cependant, ces auteurs ne répondent pas à la question des raisons de l'apparition de cet attracteur. Darcheville (Darcheville, 1998, Boyer and Darcheville, 1999) émet l'hypothèse que l'objet atteint est le renforçateur de ce mouvement. La valeur renforçatrice des objets (par exemple, des objets intéressants pour l'enfant) s'acquiert durant cette même période. On sait que les renforçateurs sélectionnent des patrons de réponses qui se stabilisent. Cette hypothèse implique que le mouvement de positionnement de la main, ou de l'atteinte, peut s'installer sans contrôle visuel du mouvement, et s'oppose donc aux conjectures piagetiennes exigeant une coordination des mouvements de la main et de l'oeil (). Les données expérimentales (Clifton and Ashmead, 1993) infirment la nécessité d'une coordination visuo-motrice.

Pour en revenir au système MAABAC, la position

pour laquelle il reçoit le renforcement peut être aisément comparée à un objet atteint de manière fortuite, et qui s'avère renforçant. Comme l'évolution du comportement de MAABAC suit celle du mouvement d'atteinte de l'enfant, on peut considérer ce système comme une simulation de l'émergence de ce comportement chez le jeune enfant par l'effet de la sélection par renforcement.

5. Discussion et perspectives

Nous avons présenté les motivations pour créer le système MAABAC et son utilisation pour simuler l'évolution de mouvements coordonnés d'un bras qui atteint une cible.

Les résultats du système MAABAC montre qu'un système non centralisé est capable de réaliser des apprentissages complexes très comparables à ceux observés chez des animaux ou chez le bébé humain. Il apprend par shaping; il montre une extinction des comportement. Bien que ce ne soit pas rapporté ici, il est aussi capable de suivre une cible mobile. Clairement, ce travail va dans le sens d'une évolution des comportements de l'animal par le biais de la sélection. Nous devons mentionner les travaux réalisés sur l'animat Darwin III (Reeke et al., 1989). Darwin III comporte également un bras multi-segment et un . Cependant, l'architecture est différente de la notre et, surtout, les motivations sont très différentes des nôtres. Darwin III repose sur la théorie d'Edelman de la sélection des groupes de neurones (TNGS ou «*Theory of Neuronal Group Selection*») (Edelman, 1987) et le but est d'implémenter les idées de la TNGS et de démontrer leur intérêt. La TNGS est une application des processus sélectionnistes de la structure des connexions synaptiques dans le cerveau d'un animal. Elle est basée sur une simulation fine de neurones réels et de réseaux de neurones existants. Darwin III repose donc sur des processus neurophysiologiques. Ce n'est clairement pas le même niveau de description que nous adoptons puisque nous ne considérons pas, actuellement, l'implémentation biologique des processus sélectionnistes. En revanche, nous nous focalisons d'avantage sur la dynamique de la sélection elle-même, ainsi que sur la dynamique de processus sélectionnistes en interaction. Par ailleurs, nous cherchons à réaliser des comportements émergent plus complexes et à évaluer leur fidélité avec les comportements d'organismes vivants. Finalement, Darwin III est composé de 50 000 neurones de 50 types différents, interconnectés par 620 000 connexions synaptiques. Ceci doit être comparé à l'architecture très simple de MAABAC.

Ceci dit, MAABAC est une simplification, presque une caricature, comparé à la diversité des contraintes d'un système physique comme le bras d'un enfant. Nous envi-

sageons donc d'intégrer, entre autres, les problèmes liés à la gravité, aux nombreux segments et muscles mis en jeux dans un mouvement ainsi que le déplacement en trois dimensions. Ce développement nous impose d'envisager des questions de divers ordres. Ces problèmes relèvent de la physique, de la bio-mécanique, de la neuro-physiologie. Ceci impose aussi d'envisager l'impact des algorithmes de contrôle.

Les développements futurs quant à la simulation par ce robot de l'émergence du comportement d'atteinte chez l'enfant suppose qu'il produit des caractéristiques du comportement infantin qui n'ont pas été envisagées dans cette simulation. Un aspect essentiel de la dynamique du mouvement d'atteinte mature est son accélération initiale puis sa décélération lorsque la main approche l'objet à atteindre. On peut imaginer que cette dynamique résulte de la nécessité de saisir l'objet. Cela suppose donc que l'on adjoigne aux deux segments existant une pince, automate autonome, comme substitut de la main. Donc contrairement au procédé habituel (Berthier, 1996) de la simulation du reaching où l'on part de la dynamique observée d'un comportement stabilisé et efficace pour envisager un système d'équations la décrivant, nous proposons de mettre en coopération des comportements d'automates sous le contrôle de la même contingence de renforcement. C'est de cette coopération sous cette contingence que doit émerger la dynamique observée.

Enfin, et compte tenu de l'explosion combinatoire des paramètres que nous aurons à fixer, nous envisageons de recourir à des méthodes évolutionnaires (algorithmes génétique) pour rechercher les optimums de ces paramètres. Cette voie de recherche présente selon nous un double intérêt. D'une part, il est très difficile de connaître à priori ou par une méthode exacte les « bonnes valeurs » des paramètres. D'autre part, les systèmes biologiques résultant eux-mêmes d'une évolution sélectionniste, les méthodes évolutionnaires paraissent parfaitement cohérentes (Kodjabachian, 1998, Parisi et al., 1991).

Remerciements

Nous tenons à remercier Francine Ciancia et Christine Demaire du Laboratoire de Neurosciences du comportement de l'Université de Lille I, pour les conseils qu'elles nous ont apportés notamment pour les questions de *neuro-physiologie*.

Cette recherche fait l'objet d'un financement par le Conseil Régional du Nord-Pas de Calais (contrat n97 53 0283).

Références

- Axelrod, R., (Ed.) (1984). *The evolution of cooperation*. Basic Book, New York, USA.
- Berthier, N. (1996). Learning to reach: A mathematical model. *Developmental Psychology*, 32:811-823.
- Boyer, C. and Darcheville, J.-C. (1999). Interaction organisation-environnement dans l'émergence de la saisie chez le nourrisson. In *Congrès National de Société Française de Psychologie*.
- Calabretta, R., Nolfi, S., Parisi, D., and Wagner, G. (1998). A case study of the evolution of modularity: Towards a bridge between evolutionary biology, artificial life, neuro- and cognitive science. In (?).
- Clifton, R. M. and Ashmead, D.H. and Clarkson, M. (1993). Is visually guided reaching in infancy a myth? *Child Development*, 64:1099-1110.
- Cuvelier, B., Preux, P., and Cambier, C. (1997). Studying adaptation with ECHO. In Husbands, P. and Harvey, I., (Eds.), *European Conference on Artificial Life (ECAL) online proceedings*, <http://www.cogs.susx.ac.uk/ecal97/present.html>.
- Darcheville, J.-C. (1998). Rapport interne, Unité de Recherche sur l'Évolution des Comportements et des Apprentissages, Université de Lille 3.
- Delepoulle, S., Preux, P., and Darcheville, J.-C. (2000). Evolution of cooperation within a behavior-based perspective: confronting nature and animats. In *Evolution artificielle 1999, Dunkerque*, Lecture Notes in Computer Science. Springer-Verlag.
- Edelman, G. (1987). *Neural Darwinism*. Basic Books.
- Forrest, S. and Jones, T. (1994). Modeling complex adaptive systems with Echo: Mechanisms of adaptation. In Stonier, R. and Yu, X., (Eds.), *Complex Systems: Mechanisms of Adaptation*, pages 3-21. IOS Press.
- Forssberg, H. (1986). Development and integration of human locomotor function. *Development and Plasticity of the Mammalian Spinal Cord* M.E. Goldberger, A. Gorio, M. Murray (eds.) *Fidia Research Series, Vol. III*.
- Hake, D. F. and Olvera, D. R. (1978). Cooperation, competition, and related social phenomena. In Catania, A. C. and Brigham, T. A., (Eds.), *Handbook of applied behavior analysis*, pages 208-245. New-York: Irvington.
- Hake, D. F. and Vukelich, R. (1972). A classification and review of cooperation procedures. *Journal of the Experimental Analysis of Behavior*, 18:333-343.
- Hake, D. F. and Vukelich, R. (1973). Analysis of the control exerted by a complex cooperation procedure. *Journal of the Experimental Analysis of Behavior*, 19:3-16.

- Hemelrijk, C. (1997). Cooperation without genes, games or cognition. In *Proceeding of European Conference on Artificial Life 1997*.
- Killingback, T., Doebeli, M., and Knowlton, N. (1999). Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. *Proceedings of the Royal Society of London*.
- Kodjabachian, J. (1998). *Développement et évolution de réseaux de neurones artificiels*. PhD thesis, Paris 6.
- Maes, P. and Brooks, R. A. (1990). Learning to coordinate behaviors. *Journal of Abnormal Social Psychology*, 52:115–119.
- Mataric, M. J. (1995). Designing and understanding adaptive group behavior. *Adaptive Behavior*, 4:51–80.
- Mataric, M. J., Williamson, M., Demiris, J., and Mohan, A. (1998). Behavior-based primitives for articulated control. In Pfeifer, R., Blumberg, B., Meyer, J.-A., and Wilson, S. W., (Eds.), *From Animals to Animats 5: Proceedings of the fifth international conference on simulation of adaptive behavior*, pages 165–170. Cambridge, MA: The MIT Press/Bradford Books.
- Mitchell, M. and Belew, R., (Eds.) (1996). *Adaptive Individuals In Evolving Population Models*. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley Publishing Company.
- Parisi, D., Nolfi, S., and Cecconi, F. (1991). Learning, behavior, and evolution. In Varela, F. and Bourgine, P., (Eds.), *Towards a practice of autonomous systems, Proc. of the First European Conference on Artificial Life (ECAL)*, pages 207–216. MIT Press.
- Ray, T. S. (1994). An evolutionary approach to synthetic biology, Zen and the art of creating life. *Artificial Life*, 1(1).
- Reeke, G., Sporns, O., and Edelman, G. (1989). Synthetic neural modelling: Comparisons of population and connectionist approaches. In Pfeifer, R., Schreter, Z., Fogelman-Soulié, F., and Steels, L., (Eds.), *Connectionism in Perspective*. Elsevier Science Publishers.
- Rosenzweig, M. R. and Leiman, A. L. (1989). *Physiological Psychology Second Edition*. Random House, Inc.
- Scheier, C. and Pfeifer, R. (1998). Exploiting embodiment for category learning. In (?), pages 32–37.
- Schmitt, D. R. (1976). Some condition affecting the choice to cooperate or compete. *Journal of the Experimental Analysis of Behavior*, 25:165–178.
- Sidowski, J. (1957). Reward and punishment in a minimal social situation. *Journal of Experimental Psychology*, 55:318–326.
- Sidowski, J. B., Wyckoff, B., and Tabor, L. (1956). The influence of reinforcement and punishment in a minimal social situation. *Journal of Abnormal Social Psychology*, 52:115–119.
- Skinner, B. F. (1938). *The behavior of organisms*. Appleton-Century Crofts. new edition available via the Skinner foundation, www.bfskinner.org.
- Spier, E. and McFarland, D. (1998). Learning to do without cognition. In (?), pages 38–47.
- Steels, L. (1994). The artificial life roots of artificial intelligence. *Artificial Life*, 1(1):75–110.
- Sutton, R. and Barto, A. (1990). Time-derivative models of pavlovian reinforcement. In Gabriel, M. and Moore, J., (Eds.), *Learning and Computational Neurosciences: Foundations of Adaptive Networks*, pages 497–537. MIT Press.
- Thelen, E. and Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge: The MIT Press.
- Touretzky, D. S. and Saksida, L. M. (1996). Skinnerbots. In Maes, M. J. M. P., Meyer, J.-A., Pollock, J., and Wilson, S. W., (Eds.), *From Animals to Animats 4: Proceedings of the fourth international conference on simulation of adaptive behavior*. Cambridge, MA: The MIT Press/Bradford Books.
- Watkins, C. (1989). *Learning from delayed rewards*. PhD thesis, King's college, Cambridge, UK.