THÈSE de DOCTORAT de L'UNIVERSITÉ PARIS 6

Spécialité : **Informatique**

Présentée par : **Benoît GIRARD**

pour obtenir le grade de
DOCTEUR de L'UNIVERSITÉ PARIS 6

Intégration de la navigation et de la sélection de l'action dans une architecture de contrôle inspirée des ganglions de la base

Soutenue le 12 septembre 2003,

devant le jury composé de :

Pr. Alain BERTHOZ	(co-directeur de thèse)	LPPA, Collège de France
Dr. Raja CHATILA	(rapporteur)	LAAS, CNRS
Pr. Jean-Michel DENIAU	(examinateur)	U114, Université Paris VI
Pr. Philippe GAUSSIER	(examinateur)	ETIS, Université de Cergy-Pontoise
Dr. Agnès GUILLOT	(co-directeur de thèse)	LIP6, Université Paris X
Pr. Eric HORLAIT	(examinateur)	LIP6, Université Paris VI
Dr. Tony PRESCOTT	(rapporteur)	ABRG, University of Sheffield

Résumé

La conception d'architectures de contrôle de robots adaptatifs autonomes nécessite de résoudre les problèmes de sélection de l'action et de navigation. La sélection de l'action concerne le choix, à chaque instant, du comportement le plus adapté afin d'assurer la survie. Ce choix dépend du contexte environnemental, de l'état interne du robot et de motivations pouvant être contradictoires. La navigation se rapporte à la locomotion, la cartographie, la localisation et la planification de chemin dans l'environnement. La mise en œuvre conjointe de ces deux capacités –pour, par exemple, exploiter la planification de chemin pour retrouver des ressources vitales– n'a été que peu abordée par les nombreux systèmes ingénieurs appliqués à la robotique autonome. Les progrès récents en neurosciences permettent de proposer des modèles des structures neurales impliquées dans l'intégration d'information spatiales pour la sélection de l'action. Chez les vertébrés, ces structures correspondent aux ganglions de la base, un ensemble de noyaux subcorticaux.

L'objectif de ce travail a été de s'inspirer de ces connaissances neurobiologiques pour élaborer l'architecture de sélection de l'action d'un robot autonome prenant en compte à la fois des informations sensorimotrices, motivationnelles et spatiales.

Dans un premier temps, nous avons adapté un modèle biomimétique de sélection de l'action déjà existant pour tester sa capacité à résoudre une tâche de survie dans une implémentation robotique. Nous avons montré, par des comparaisons avec un système de sélection de type « winner-takes-all », que ses propriétés dynamiques lui permettent de limiter les oscillations comportementales, de maintenir ses variables internes à un niveau plus élevé et de limiter sa consommation d'énergie.

Dans un deuxième temps, nous nous sommes inspirés des rôles distincts des circuits dorsaux –sélection de l'action– et ventraux –intégration de la navigation– des ganglions de la base pour élaborer une architecture interfaçant ce modèle de sélection de l'action avec deux stratégies de navigation : approche d'objets et planification topologique. Nous l'avons testée sur un robot simulé réalisant une tâche de survie similaire à la précédente. Le robot s'est avéré capable d'utiliser la planification pour rejoindre des ressources distantes, d'utiliser de façon complémentaire l'approche d'objets pour exploiter les ressources inconnues, d'adapter son comportement à la disparition de ressources, à son état interne et aux configurations environnementales, et enfin de survivre dans un environnement complexe réunissant l'ensemble des situations préalablement testées.

Nous concluons que les circuits des ganglions de la base modélisés ont permis d'obtenir un système robuste d'interface de la sélection de l'action et de la navigation pour une architecture de contrôle de robot autonome. Cependant, des connaissances supplémentaires en neurobiologie seraient nécessaires pour affiner la plausibilité du modèle proposé. De plus, l'intégration de capacités d'apprentissage par renforcement –qui mettent également en jeu les ganglions de la base– s'avère indispensable pour améliorer l'adaptativité de notre modèle.

Mots-clés: animat, biorobotique, ganglions de la base, modèle computationnel, navigation, sélection de l'action

Abstract

The functions of action selection and navigation are essential for the design of adaptive robots. Action selection concerns the choice, at all time, of the most appropriate behaviour that maintains survival. This choice depends on the environmental context, and on the internal states and motivations of the robot. Navigation deals with locomotion, mapping, localization and path planning. The integration of both these abilities is crucial for an autonomous robot, for instance to retrieve hidden resources by using path planning. Up to now, few engineering models have focused on interfacing them. Recently, models of the basal ganglia –vertebrate neural structures, implicated in the integration of spatial information for action selection, have been proposed.

The aim of our work has been to build from this neurobiological knowledge an architecture for action selection, which is able to integrate sensorimotor, motivational and spatial information. We have first adapted an existing biomimetic model of action selection and tested its ability to solve a survival task in a robotic implementation. The comparison with a simple « winner-takes-all » mechanism has demonstrated that the dynamical properties of the model provides adaptive capacities, e.g. limitation of behavioural dithering, maintenance of internal variables at higher levels and reduction of energy consumption

Drawing inspiration from the distinct roles of the dorsal and ventral circuits of the basal ganglia –resp. action selection and navigation integration, we have then elaborated an architecture interfacing this model with two navigation strategies : object approach and topological navigation . A simulated robot, tested in a similar survival task, has been able to use both path planning and object approach to reach distant resources and exploit unknown supply, to cope with various internal states and environmental configurations, and to survive in a large environment where all the previous abilities had to be exhibited.

We came to the conclusion that the basal ganglia circuits would provide a robust system interfacing action selection and navigation for autonomous robots. However, some complementary neurobiological knowledge would be necessary to refine the biological plausibility of our model. Adding reinforcement learning –processes that implicate the basal ganglia– is necessary to enhance the adaptivity of our architecture.

Keywords: animat, biorobotics, basal ganglia, computational model, navigation, action selection

Remerciements

Je remercie les membres du Jury d'avoir accepté de m'honorer de leur présence lors de la soutenance de cette thèse.

Tout d'abord Messieurs Chatila, Gaussier et Horlait qui ont accepté de faire partie de ce jury malgré leurs nombreuses charges et contraintes professionnelles.

Je remercie également Monsieur Jean-Michel Deniau, qui, avec Madame Anne-Marie Thierry, s'est montré patient et disponible pour répondre à mes naïves questions sur la neurobiologie du rat.

Many thanks to you, Tony Prescott, for accepting the tough charge of reviewing a thesis written in french. The interactions with Kevin Gurney, Peter Redgrave and you were very fruitful, it was a pleasure to collaborate with the three of you : the fourth chapter of this work owes much to the discussions we had in November 2002. Tell Peter I forgive him for hating modelers.

Enfin, je remercie naturellement Madame Agnès Guillot d'avoir accepté la charge de diriger ma thèse au sein de l'AnimatLab et Monsieur Alain Berthoz de s'être intéressé à nos premières réalisations de « roboticiens », au point de proposer une co-direction, venant y ajouter la touche du « biologiste ». Les interactions avec mes deux directeurs furent toujours stimulantes et pleines d'enseignements.

Je remercie chaleureusement Jean-Arcady Meyer et Agnès Guillot, récemment rejoints par Olivier Sigaud, pour m'avoir initié aux joies de l'approche animat et pour avoir réuni autour d'eux une équipe sympathique et soudée où il a fait si bon travailler, trois ans durant. Merci donc à Gildas Bayard pour ses bruitages, à Vincent Cuzin pour son gros disque dur, à David Filliat pour son indéfectible calme, à Fabien Flacher pour ses mythiques coups de grogne, à Thomas Degris pour CS, à Stéphane Doncieux pour sa gentillesse, à Pierre Gérard pour sa bonne humeur et sa générosité, à Thierry Gourdin pour son impeccable bouc et le sourire qui toujours le surmonte, à Stéphane Gourichon pour son mode verbose et ses conseils mathématiques avisés, à Mehdi Khamassi pour son enthousiasme et ses grandioses visions d'avenir, à Loïc Lachèze pour son dévouement, à Sébastien Laithier pour son temps et sa vieille carte vidéo, à Gabriel Robert pour et malgré sa nature de troll et enfin à tous pour m'avoir accordé leur amitié.

Mes visites au LPPA furent sûrement trop rares, je remercie Sidney Wiener pour m'avoir mis sur les rails avec ses conseils de lecture avisés et Angelo Arleo pour les multiples explications de son modèle qu'il a bien voulu me prodiguer.

Les travaux présentés dans ce rapport n'auraient sûrement pas été menés à bien sans l'aide technique de Vincent Cuzin pour la communication IR du robot Lego et la détection de bug dans le noyau de LegOS, de Sébastien Laithier pour son codage XML amusant des calculs de salience et son temps consacré à des expériences, de David Filliat pour son service après-vente

irréprochable et enfin de Bernard Girard pour ses mises à jour en statistiques. Je les remercie donc à nouveau et en profite pour souhaiter du courage à Loïc et Mehdi qui hériteront du code résultant de ces interactions.

Je remercie l'ensemble du personnel administratif du LIP6 qui m'a permis, en prenant en charge le fonctionnement de ma thèse, de me concentrer sur son contenu, en particulier Nicole Bohelay, Chantale Darin, Jacqueline Le Baquer, Ghislaine Mary et Christophe Bouder pour leur efficacité et leur caractère jovial.

Ma vocation pour la recherche n'est pas récente. Je la dois tout d'abord à une famille pleine de chercheurs sympathiques : mes parents Colette et Michel, mes oncles et tantes Bernard, Yvonne, Françoise et Gérard. L'envie de rendre possible ce que Monique Lebailly m'avait fait découvrir dans la science-fiction m'aura guidé vers l'IA et les robots autonomes. M et Mme Monchamp, enseignants en biologie de grande qualité, m'auront d'une part fait découvrir le sens d'une démarche scientifique rigoureuse et d'autre part fourni des connaissances de base qui me furent utiles jusque durant cette thèse. Ils auront su éveiller mon intérêt pour la biologie. Ma découverte de la vie de laboratoire de l'interieur eut lieu auprès du formidable duo composé de Lionel Bresson et Denis Gratias, j'en souhaite autant à tout apprenti chercheur.

Toujours sur un plan personnel, ma famille et de nombreux amis fidèles ont été présents, en particulier lorsqu'il semblait nécessaire de me changer les idées, je remercie donc Florent et Amélie Castelnérac, Eddy et Blandine Collin, Sébastien Cagnoli, Cyrille Boulanger, Delphine Bourguignon, Lionel Bringuier, Stéphane Gigandet, Régis Lorient, Catherine Mussillon, Jeff Portet, Sébastien Ribaute et sûrement d'autres que j'oublie et qui m'en voudront –à raison– très longtemps.

Je remercie Charles Abelé, David Fusco-Vigné et l'ensemble des membres de l'APA 14ème pour leur pratique d'un aïkido épanouissant et pour s'être efforcés de me conserver un corps sain à défaut d'avoir pu faire quoi que ce soit pour mon esprit.

J'exprime aussi ma reconnaissance aux concepteurs de BZflag, Super Puzzle Fighter Turbo II fx, XBlast et Counter Strike, pour avoir participé de la bonne ambiance qui règne à l'Animat-Lab et pour les froncements de sourcil si amusants qu'ils n'ont pas manqué de faire naître chez Olivier.

La musique surpasse toute autre forme de dopage, je remercie donc tout d'abord Infectious Grooves, Jimmy Hendrix, The Lords of Acid, Rammstein, Rhapsody, Suicidal Tendencies et Frank Zappa pour m'avoir stimulé lors des tâches de programmation parfois si laborieuses et ingrates, puis Garbage, Alain Bashung, Morcheeba, Purcell et Heitor Villa-Lobos pour m'avoir accompagnés lors de la rédaction de ce rapport.

Enfin, ayant gardé le meilleur pour la fin, c'est sans grande originalité mais avec grande sincérité que je remercie ma femme Valérie pour son intérêt pour mon travail, ses questions toujours pertinentes et surtout son soutien lors des moments difficiles qui émaillent nécessairement la rédaction d'un rapport de thèse. Bon vent pour ta propre thèse !

Table des matières

Introduction

Les animaux sont capables d'assurer leur survie de façon totalement autonome dans des environnements complexes, dynamiques et bien souvent hostiles. Cette survie dépend entre autres de leur capacité à exhiber des comportements adaptés à la fois aux besoins de leur métabolisme et au contexte environnemental. Ils sont capables de sélectionner à chaque instant, parmi un ensemble d'objectifs contradictoires, celui qu'il convient de mener à bien et la manière d'y parvenir. Cette capacité est désignée sous le nom de *sélection de l'action*. La mise en œuvre de ces comportements, en particulier chez les vertébrés, nécessite l'utilisation de stratégies de *navigation* élaborées, permettant par exemple de trouver dans l'environnement les ressources nécessaires et de les rapporter au nid tout en évitant les prédateurs. La navigation désigne toute aptitude d'un animat à se repérer, s'orienter et se déplacer dans son environnement.

Ces propriétés de sélection de l'action et de navigation sont également nécessaires au bon fonctionnement d'un robot autonome. Les progrès récents dans la compréhension des bases neurales de ces capacités permettent d'envisager leur modélisation computationnelle et leur intégration sur plateforme robotique. S'inspirer ainsi des sciences du vivant pour concevoir un système artificiel adaptatif relève de « l'approche animat ».

Dans les paragraphes suivants, nous allons introduire cette approche qui s'intéresse à la conception de robots autonomes en s'inspirant des propriétés animales. Nous allons également évoquer la démarche biomimétique qui, au sein de l'approche animat, se consacre à la conception de modèles computationnels de systèmes nerveux animaux. Enfin, nous allons présenter les objectifs de cette thèse, qui se situe dans le cadre du projet Psikharpax fédérant de nombreuses problématiques de l'approche animat.

1 Approche animat

L'*Approche Animat* vise à concevoir des animaux artificiels (animats) simulés ou robotiques, au fonctionnement inspiré des animaux. Son objectif est de comprendre les mécanismes d'autonomie et d'adaptation des animaux, puis de les importer dans des artefacts capables, eux aussi, de s'adapter et d'assurer leur mission dans un environnement dynamique imprévisible (Meyer et Guillot, 1991; Wilson, 1991; Meyer, 1996).

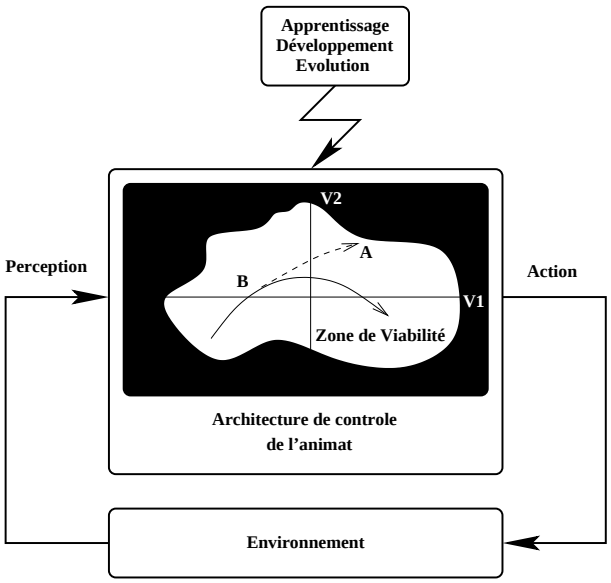

FIG. *1: Le comportement d'un animat peut être qualifié d'adaptatif tant que son architecture de contrôle permet de maintenir ses variables essentielles (ici V1 et V2) dans leur zone de viabilité. Ici, une action correctrice a été accomplie au point B de façon à éviter de quitter la zone de viabilité au point A. Dans la mesure où cette architecture de contrôle sert aussi à choisir les buts successifs que l'animat cherche à atteindre en arbitrant entre des buts conflictuels, elle joue le rôle d'un système motivationnel. L'organisation de l'architecture de contrôle peut être modifiée par des processus de développement, d'apprentissage ou d'évolution (d'après Meyer et Guillot, 1991).*

Cette voie de recherche est née d'un certain nombre de limitations de l'intelligence artificielle classique (IAC), résumées par Dreyfus (1972) et Guillot et Meyer (2003) en ces quelques points :

- le système de décision y est considéré comme isolé du monde, la notion d'enveloppe corporelle et la résolution des problèmes d'interaction avec le monde réel sont considérés comme négligeables,
- le concepteur doit conséquemment prétraiter l'information issue du monde réel pour pouvoir fournir ses entrées au système, puis en interpréter les sorties,
- il doit également envisager de manière exhaustive l'ensemble des situations susceptibles d'être rencontrées lors de la conception du système, ce qui n'est guère compatible avec l'interaction avec l'environnement réel.

L'IAC se montre donc performante pour la réalisation de systèmes qui « raisonnent », mais moins efficace dans le cadre de systèmes qui « se comportent » tels que les robots, à moins qu'ils n'évoluent en environnements contrôlés (robots industriels).

L'approche animat tire son inspiration des sciences du vivant. Multidisciplinaire par définition, elle se place au carrefour de différentes disciplines : informatique, robotique, sciences cognitives, éthologie, biologie en général et neurosciences en particulier. Cet intérêt pour le vivant est issu de la simple constatation des capacités d'adaptation et d'autonomie développées par les êtres vivants dans leur milieu naturel, bien supérieures à celles des robots actuels.

La notion d'autonomie telle qu'elle est entendue dans le cadre de cette approche diffère de celle des ingénieurs et des industriels (Alami *et al.*, 1998). Elle correspond en effet au fait que le système peut de lui-même agir de telle façon qu'il conserve toujours ses variables « essentielles » dans une zone de viabilité (Ashby, 1952), dans des environnements dynamiques et imprévisibles, soumis parfois à des contraintes déclaratives (mission à accomplir) mais toujours libre du choix de ses procédures (moyens) mises en œuvre pour s'y soumettre (fig. 1). On comprend que cette dernière possibilité puisse susciter quelques réticences chez les industriels. Ainsi que le résume fort bien (Keijzer, 1998) :

« A key issue of adaptive behavior is not merely to achieve distal goals, but to achieve these goals under always varying proximal circumstances. »

Les fondations de cette approche étant établies, une méthodologie spécifique en découle :

- les systèmes élaborés ont pour destinée d'être *situés*, embarqués sur des plateformes robotiques évoluant dans le monde réel, ils doivent donc être complètement spécifiés, depuis la perception jusqu'à l'action en passant par l'architecture de contrôle,
- ils sont conçus par une ingénierie inverse appliquée aux systèmes naturels (*« when the 'devices' were already 'designed' and 'built' by nature - by evolution - and we have to fi gure out how they work, how they can do what they can do »* –(Dawkins, 1995)), dans une approche *bottom-up*, où l'on commence par reproduire les comportements les plus simples pour les utiliser ensuite comme constituants de base de comportements plus complexes,
- enfin, le degré de réductionnisme adopté dans la modélisation est varié : il peut être très proche de la réalité observée, comme dans la modélisation du système de vision de la mouche de (Franceschini *et al.*, 1992), plus éloigné, dans les très nombreux modèles utilisant comme composant de base les neurones artificiels –si simples comparés à leur homologues naturels–, voire lointain, lorsque l'on s'intéresse notamment aux méthodes d'évolution artificielle dans lesquelles un darwinisme primaire est mis en œuvre.

Enfin, la perspective de l'approche animat est double : il s'agit à la fois, d'un point de vue fondamental, de chercher à comprendre les mécanismes adaptatifs du vivant et, d'un point de vue appliqué, d'attribuer des capacités d'autonomie aux systèmes artificiels (Webb, 2001;

Meyer et Guillot, 1994; Guillot et Meyer, 2000; Guillot et Meyer, 2001). C'est de ce flux bidirectionnel que peuvent naître de réelles interactions avec les sciences du vivant.

1.1 De l'animat à l'animal

L'approche animat aspire donc non seulement à doter les robots autonomes des capacités des animaux (voir Bar-Cohen et Breazeal, 2003 pour une revue), mais également à nourrir en retour les sciences du vivant (Webb, 2001).

En intégrant plusieurs mécanismes étudiés isolément dans des systèmes simulés ou robotiques, elle permet, par exemple, de tester des hypothèses spécifiques –comme celles concernant les règles de navigation guidant le comportement des fourmis (Lambrinos *et al.*, 2000), des abeilles (Srinivasan *et al.*, 1999) ou des rats (Burgess *et al.*, 1997)– ou des hypothèses plus générales, comme de mettre en œuvre la « loi de la parcimonie » (Morgan, 1894), en se demandant dans quelle mesure l'interaction de mécanismes réactifs peut conduire à des comportements dits cognitifs (Brooks, 1991) et pour quels problèmes environnementaux ces derniers sont réellement nécessaires.

Enfin elle présente des commodités méthodologiques qui la rendent complémentaire des expérimentations animales, avec un contrôle plus rigoureux du système étudié (pas d'influence du mode d'élevage ou d'éducation du robot sur son comportement présent, pas de comportements « parasites » produits par des mécanismes non-étudiés) et un accès à des variables cachées, comme les variables internes d'un animal.

De même, par des modèles incrémentaux, elle peut déterminer les rôles respectifs des divers mécanismes supposés être impliqués dans la production d'un comportement. L'équipe de Webb (Webb, 1995; Webb et Scutt, 2000), par exemple, s'est tout d'abord intéressée à l'étude d'un mécanisme neuronal suffisant pour expliquer à la fois le comportement d'approche de la femelle et la sélectivité du chant chez le grillon, puis a intégré un système auditif périphérique expliquant des caractéristiques spatio-temporelles de son comportement et enfin a raffiné le modèle de neurones artificiels afin d'en perfectionner les aspects temporels.

1.2 Validation des modèles

Il reste cependant que l'application de cette même méthodologie est susceptible de limiter la portée de sa contribution aux sciences du vivant (Chatila, 2002; Webb et Consi, 2001; Florian, 2003).

En effet, les simplifications qu'implique la conception de modèles computationnels ne sont pas toujours compatibles avec le détail des mécanismes biologiques étudiés. Or, il s'avère difficile d'évaluer où se situe la frontière pertinente entre modélisation fidèle et trop haut niveau d'abstraction.

L'évaluation et la validation des modèles est difficile. Elles sont trop souvent qualitatives (« l'animat se comporte comme un animal ») donc subjectives et arbitraires. Le développement d'évaluations quantitatives et comparatives est indispensable. Il peut s'agir, dans un premier temps, de comparer entre eux plusieurs modèles computationnels, bioinspirés ou non, ou de comparer incrémentalement les versions successives d'un modèle. Encore faut-il que l'environnement de test soit pertinent. En effet, les mécanismes biologiques sont supposés avoir évolué pour résoudre des problèmes donnés, dans des environnements donnés. Souvent le biologiste ignore à quel problème environnemental précis telle ou telle organisation anatomique ou nerveuse apporte une solution. Si l'environnement du robot, équipé d'un modèle computationnel copié sur ces mécanismes, ne lui fournit pas les mêmes contraintes, le rôle de ces mécanismes sera mal interprété ou restera ignoré.

Une solution –en général inaccessible à cause des limitations techniques– est la comparaison de systèmes artificiels et de systèmes vivants, testés rigoureusement dans les mêmes conditions expérimentales.

1.3 Démarche biomimétique et biorobotique

Au sein de l'approche animat, la démarche biomimétique se place en position d'interaction avec les neurosciences. Elle se situe à un degré de réductionnisme intermédiaire, adoptant couramment le modèle des réseaux de neurones artificiels. Elle donne lieu, d'une part, à des travaux de simulation (Baldassare, 2003), nécessaires pour tester rapidement des hypothèses et préparer la réalisation de robots en s'affranchissant des problèmes techniques, et, d'autre part, à des implémentations robotiques (Webb et Consi, 2001) qui sont capitales pour prendre en compte l'imprédictibilité intrinsèque d'un environnement réel.

C'est dans le cadre général de l'approche animat, et dans celui plus restreint de la démarche biomimétique, que se place notre travail.

2 Projet Psikharpax

Ce travail de thèse est partie intégrante du projet Psikharpax. Il s'agit un projet ROBEA et LIP6 qui réunit trois équipes académiques françaises –l'AnimatLab du Laboratoire d'Informatique de Paris 6 (LIP6), le Laboratoire de Physiologie de la Perception et de l'Action (LPPA) du Collège de France, et le Laboratoire d'Informatique et de Microélectronique de Montpellier (LIRMM)– ainsi que deux équipes étrangères –le Laboratoire de Calcul Neuromimétique de l'Ecole Polytechnique Fédérale de Lausanne et l'Adaptive Behaviour Research Group de l'Université de Sheffield. Il implique également deux partenaires industriels, les sociétés Wany et BEV.

Il vise à la synthèse d'un « rat artificiel » implémentant des mécanismes adaptatifs mis en évidence sur le rat réel, mécanismes qui permettent à cet animal de survivre dans un environnement inconnu et changeant. L'intégration cohérente de ces mécanismes doit à terme être démontrée sur une plate-forme robotique.

Les objectifs de ce projet sont, d'un côté, de tester la cohérence et la complétude des connaissances sur le fonctionnement du système nerveux du rat et sur les mécanismes qui concourent à ses capacités adaptatives et, d'un autre côté, de mettre au point un robot adaptatif capable d'autonomie dans le choix de ses buts et de ses actions (Meyer, 2002; Guillot et Meyer, 2002).

2.1 Une démarche intégrative

Bien que les travaux de l'approche animat aient fait état de nombreux progrès dans la conception et la mise au point de senseurs ou d'effecteurs artificiels inspirés de ceux des animaux, il y a encore peu de travaux portant sur les architectures de contrôle complètes de robots ou d'agents virtuels.

Lorsque de tels travaux existent, ils sont essentiellement inspirés des systèmes nerveux des invertébrés (Beer et Chiel, 1991). Les rares qui se soient inspirés des comportements et des capacités adaptatives des vertébrés sont généralement centrés sur des comportements particuliers (par exemple, la locomotion de la salamandre (Ijspeert, 2001) ou de l'humain (Bongard et Paul, 2000)).

L'originalité du projet Psikharpax réside dans l'intégration de multiples mécanismes de contrôle de haut niveau coordonnant ainsi des comportements aussi nombreux, différents et éventuellement incompatibles que le repos, la locomotion, l'alimentation, ou l'évitement de prédateurs.

2.2 Pourquoi le rat ?

Dans le cadre d'une intégration biomimétique de plusieurs fonctionnalités étudiées séparément, les connaissances sur les mécanismes correspondants doivent être suffisamment nombreuses et précises. Le rat est l'un des animaux de laboratoire les plus étudiés par les comportementalistes (psycho- et neurophysiologistes, psycho- et neuropharmacologistes). La souris est certes génétiquement plus connue que le rat, mais sa variabilité comportementale est plus importante. De nombreuses connaissances ont été révélées chez le rat, comme par exemple différents types d'apprentissage (apprentissage latent (Tolman et Honzik, 1930) ; apprentissages skinneriens (Skinner, 1938)), sur la cognition spatiale (cellules de lieux, de direction de la tête, codage allocentriques et égocentriques, carte cognitive, (e. g. McNaughton *et al.*, 1993) ou sur l'effet collectif du stress (Calhoun, 1962).

2.3 Organisation du projet

Le projet Psikharpax est conduit en plusieurs étapes, dont un certain nombre sont menées en parallèle (voir fig. 2), donnant lieu à des tests en simulation ou sur des robots du commerce (Lego, Pioneer, Pekee) :

- test des architectures de contrôle pour la sélection de l'action et la navigation isolément,
- interfaçage de ces systèmes,
- intégration des processus d'apprentissage dans la sélection de l'action, test des systèmes de vision, audition, toucher, proprioception biomimétiques du rat.

Les travaux correspondant sont destinés à être portés à terme sur le robot prototype Psikharpax, actuellement en construction.

Afin de faciliter son évaluation finale, un certain nombre de fonctionnalités ont été fixées comme objectifs du projet (Meyer, 2002).

Psikharpax devra :

(i) utiliser ses réflexes de base pour se déplacer dans son environnement et éviter les obstacles qui s'y trouvent,

(ii) utiliser des stratégies efficaces d'exploration de l'environnement et de détection des amers rencontrés,

(iii) fusionner les informations visuelles acquises sur ces amers avec les informations proprioceptives concommitantes, afin de pondérer leurs influences respectives en fonction du contexte et d'élaborer une « carte cognitive » de son environnement,

(iv) utiliser cette carte pour se positionner lui-même et pour localiser les endroits où des récompenses ou des punitions ont été reçues,

(v) utiliser un système motivationnel pour sélectionner le but courant à satisfaire,

(vi) choisir la stratégie de navigation la plus adaptée pour rejoindre un lieu où le but courant peut être satisfait, selon que ce but est directement visible ou qu'il est mémorisé dans la carte cognitive

(vii) contrôler son bilan énergétique, notamment par une alternance de périodes d'activité et de repos.

3 Objectif de notre travail

Notre travail porte, d'une part, sur la sélection de l'action pour les robots autonomes et, en particulier, les *modèles biomimétiques des structures nerveuses* supposées assurer cette sélection chez le rat : les ganglions de la base, groupe de noyaux sub-corticaux. De tels modèles computationnels ont déjà été élaborés. Nous faisons l'hypothèse qu'ils sont susceptibles d'ap-

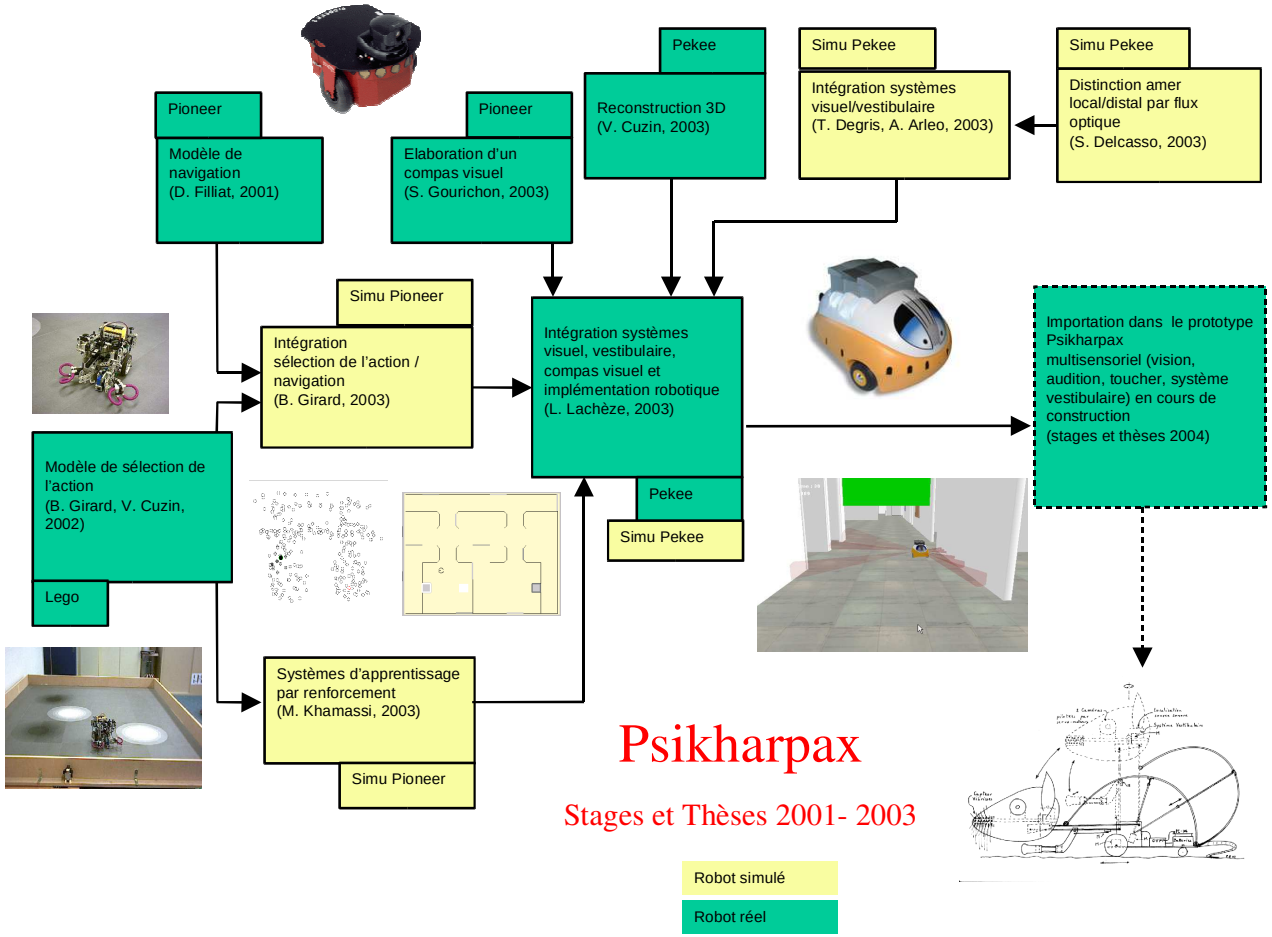

FIG. 2: *Organigramme du projet Psikharpax.*

porter à un robot certaines capacités espérées chez Psikharpax (notamment i, v et vii).

Il porte, d'autre part, sur l'*interfaçage* d'un système de navigation avec un système de sélection de l'action. De très rares modèles réalisent cette interface. A ce jour, aucun ne s'inspire des études neurobiologiques proposant des hypothèses quant à la façon dont certains circuits des ganglions de la base intègrent des informations spatiales. Nous supposons que le modèle que nous allons élaborer sur ce principe va assurer à Psikharpax –et à d'autres systèmes artificiels– plus d'adaptabilité et d'autonomie, lui permettant notamment la réalisation des capacités restantes (ii, iii, iv et vi). Nous espérons également que, par la mise en œuvre de mécanismes bioinspirés interagissant dans un système situé, ce travail puisse contribuer utilement à l'enrichissement des connaissances biologiques.

4 Plan

Nous commencerons dans le premier chapitre par donner notre définition de la sélection de l'action, car selon les domaines et les travaux, différentes capacités peuvent recouvrir ce même vocable. Un bref état de l'art des modèles computationnels de sélection de l'action chez l'animal et le robot aura pour objectif de résumer les principaux mécanismes envisagés et de mettre en évidence le rôle ambigu de la navigation dans de tels modèles.

Dans le second chapitre, nous décrirons les connaissances actuelles concernant les ganglions de la base, afin de cerner leur rôle dans le contrôle moteur et la résolution de tâches cognitives, et de définir les éléments à partir desquels les modéliser. Nous compléterons ces informations par une revue des divers modèles computationnels des ganglions de la base, qu'ils fassent ou non référence à la sélection de l'action telle que nous l'entendons.

Dans le troisième chapitre, nous décrirons le modèle computationnel que nous avons choisi parmi les modèles existants (Gurney *et al.*, 2001a) pour qu'il puisse être implémenté dans un robot réel (Lego Mindstorms ©) réalisant une tâche de survie. L'adaptation que nous en avons faite ainsi que les résultats obtenus lors de tests expérimentaux y seront également exposés.

Dans le quatrième chapitre, nous décrirons notre extension de ce modèle à la gestion de deux circuits des ganglions de la base permettant l'intégration de stratégies de navigation dans diverses tâches de sélection de l'action. Un environnement réunissant tous les problèmes précédents servira finalement à estimer, en terme de survie, les progrès apportés par le modèle.

Une discussion finale résumera dans le dernier chapitre les contributions et limitations de ce travail, ainsi que les perspectives envisagées, en terme de modélisation computationnelle de

structures nerveuses impliquées.

Chapitre 1

La sélection de l'action

Nous commencerons par définir la notion de sélection de l'action, préciser ce qui est entendu par action et évoquer le problème de l'évaluation de la qualité de la sélection à travers celui de l'optimalité de la sélection. Nous présenterons ensuite brièvement les modèles computationnels de sélection de l'action, en nous intéressant tout particulièrement, pour rester dans le cadre de notre travail, à leur éventuelle conception biomimétique puis à leur lien avec la tâche de navigation.

1.1 Définitions

1.1.1 Sélection de l'action

Le concept de « sélection de l'action » est employé dans de nombreux domaines où il a des acceptions variées. Dans le cadre de l'intelligence artificielle et de la robotique classiques, la sélection de l'action est en général associée au contrôle bas niveau de la trajectoire (locomotion ou mouvements de membres articulés) et à la planification (résolution de problèmes). En économie et en psychologie, il s'agit de la modélisation de la prise de décision. La sélection de l'action s'intéresse alors en général à la résolution d'une seule tâche, donc aux choix successifs d'actions permettant d'atteindre un seul but fixé. Elle possède alors une condition d'arrêt qui est la résolution de cet objectif.

La sélection de l'action, telle qu'elle est considérée dans le cadre de l'approche animat (Meyer et Guillot, 1991; Tyrrell, 1993a), mais également dans les travaux traitant de réalité virtuelle et de jeux vidéo (Blumberg, 1994; Girard *et al.*, 2001), est issue de l'éthologie (McFarland, 1977). Elle considère un animat (animal, robot, créature virtuelle) soumis à des objectifs concurrents et conflictuels. La sélection de l'action consiste au choix, à tout instant, de l'action la plus pertinente à mettre en œuvre afin de satisfaire au mieux ces objectifs, dans le but d'assurer sa

propre survie/son bon fonctionnement et éventuellement, dans le cas d'un robot, de réaliser les tâches pour laquelle il a été conçu. Ce choix est dépendant à la fois de l'état interne de l'animat et de celui de l'environnement, qu'il s'agisse de la seule part perceptible à un instant donné ou de zones précédemment mémorisées. La sélection de l'action n'a alors pas de condition d'arrêt, aucun objectif ne peut être résolu de façon définitive (un animal ne saurait, par exemple, se contenter de se nourrir une seule fois). Il s'agit là d'un processus actif pendant toute la durée de vie de l'animat.

Selon les mécanismes proposés par les chercheurs, la sélection de l'action peut être segmentée en diverses étapes explicites : sélection de l'objectif prioritaire, puis sélection de l'action permettant la réalisation de cet objectif et enfin mise en œuvre des commandes motrices correspondantes. Au contraire, il a aussi été proposé que ces trois étapes soient fondues en un seul processus prenant en compte toutes les informations internes et externes à l'animat et générant directement ces commandes motrices. La définition du problème de sélection de l'action ne fait pas d'hypothèse a priori concernant l'usage de l'une ou l'autre des approches permettant de le résoudre, leur point commun étant d'élucider les mécanismes d'arbitrage qui conduisent à ce que McFarland et Sibly (1975) ont appelé « the behavioral final common path », niveau de contrôle le plus bas où s'expriment les actions de l'animal.

1.1.2 Action

La définition de la sélection de l'action pèche par une imprécision quant à la définition des actions parmi lesquelles le choix opère. On peut en effet considérer, dans un premier temps, une définition issue de l'éthologie : les actions sont l'ensemble des mouvements, des vocalisations, des postures corporelles et des changements d'odeur ou de couleur (Immelmann, 1980). Cependant, la segmentation du flux comportemental continu en un ensemble d'actions discrètes pose le problème de la granularité à adopter.

En effet, les nombreux systèmes de sélection de l'action issus de l'approche animat adoptent des répertoires d'actions qui ne sont pas homogènes d'un système à l'autre. Certains considèrent des actions très élémentaires (comme de progresser d'un pas dans une direction) (Werner, 1994), d'autres des actions intégrées (comme s'approcher de la nourriture) qui nécessitent la mise en œuvre de programmes spécifiques à la résolution de leur tâche (Maes, 1989; Blumberg, 1996; Bryson, 2000). Enfin, certains auteurs refusent la notion de répertoires d'actions discrètes et proposent des systèmes contrôlant directement les moteurs, de façon totalement continue (Seth, 1998).

On notera même que, dans certains cas, le répertoire d'actions d'un même mécanisme contient des actions de granularité différente. Ainsi, Tyrrell (1993a) étudie un mécanisme de sélection de l'action doté tout à la fois de comportements locomoteurs d'assez bas niveau (se

déplacer vers l'un des quatre points cardinaux) et d'autres beaucoup plus complexes (exécuter une parade nuptiale).

L'objet de nos travaux n'est pas d'approfondir les considérations théoriques concernant la notion d'action. Notre parti-pris sera de supposer l'existence de répertoires d'actions élémentaires et de proposer des répertoires d'actions de granularité homogène.

1.1.3 Optimalité de la sélection

Un certain nombre de points spécifiques font que la sélection de l'action est un problème difficile à résoudre (Pirjanian, 1997) :

- L'animat est doté de senseurs imparfaits : ils sont bruités, peu précis, de portée limitée, etc. Les décisions doivent donc être prises malgré une connaissance imprécise de l'environnement.
- L'animat n'a également qu'une connaissance incomplète de son environnement, limitée à ce qu'il a déjà exploré. Il est donc difficile d'utiliser des techniques de planification pour résoudre la tâche. L'utilisation de la planification en sélection de l'action passe nécessairement par une phase préalable d'exploration et par la construction d'un modèle interne de l'environnement à partir duquel raisonner. Durant cette phase d'exploration, l'animat doit pouvoir assurer sa survie par d'autres méthodes.
- L'environnement est en général dynamique et imprévisible, en particulier pour les expériences robotiques. L'animat ne peut prédire de façon certaine son évolution, ce qui limite encore le recours à des méthodes de planification.
- Les actuateurs de l'animat sont imparfaits : l'animat ne peut présumer d'une exécution parfaite des ordres transmis. Il doit donc être capable d'actions correctives.
- Enfin, la tâche de sélection de l'action est sous la contrainte d'un temps limité. Ceci concerne tout d'abord une échelle de temps globale, car l'animat tend à mourir/se décharger si aucune action permettant la survie n'est entreprise, et aussi à une échelle de temps beaucoup plus restreinte, car l'opération dans l'environnement réel implique des temps de calculs et de réaction courts.

Par conséquent, la définition d'un critère d'optimalité pour la sélection de l'action est difficile à établir. La littérature sur le sujet s'est plutôt intéressée jusqu'ici à la réalisation de systèmes « satisfaisants » ou « suffisamment efficaces » pour permettre la survie de l'animat et la réalisation de son éventuelle tâche.

1.1.4 Evaluation

Ainsi que cela a été également évoqué en introduction concernant l'ensemble de l'approche animat, ce type d'imprécision sur les critères d'évaluation des systèmes proposés pose des pro-

blèmes méthodologiques. A défaut de la mise en place d'une mesure normative de l'efficacité d'un système de sélection de l'action, un certain nombre de propriétés devant être exhibées par un tel système ont été accumulées au fil des travaux (Maes, 1989; Snaith et Holland, 1991; Tyrrell, 1993a; Werner, 1994; Redgrave *et al.*, 1999a). La définition des fonctionnalités attendues de Psikharpax, proposée en introduction, relève de cette approche. Elle présente l'inconvénient de ne concerner que des propriétés isolées et parfois même partiellement contradictoires. Ainsi, on attend en général d'un système de sélection de l'action qu'il soit capable de :

- « persistance », c'est-à-dire la capacité à mener une action à bien avant de se laisser interrompre par une autre, afin d'éviter de coûteuses oscillations comportementales. Si par exemple l'animat a faim et soif, et qu'il est situé entre un point d'eau et un morceau de nourriture, il doit être capable de choisir l'une des deux ressources, de l'atteindre et de la consommer avant de s'intéresser à l'autre, plutôt que d'hésiter en approchant successivement l'une puis l'autre, sans jamais en atteindre aucune.
- « opportunisme », c'est-à-dire la capacité à interrompre une action lorsque l'occasion se présente d'en réaliser une autre sans que le coût énergétique de l'interruption soit trop grand. Si par exemple un animat qui a faim et modérément soif, alors qu'il s'approche d'une ressource de nourriture, passe à proximité d'un point d'eau, il doit pouvoir faire un détour pour boire si cela ne met pas sa survie en danger.

Ces critères ne définissent pas clairement quand une situation relève de la persistance ou de l'opportunisme. Ce flou, qui est valable pour l'ensemble de la liste des propriétés, tend à générer une évaluation trop subjective.

De plus, compte tenu de la grande complexité des systèmes de sélection de l'action et des situations spécifiques dans lesquels ils sont testés d'un travail à l'autre, les comparaisons entre systèmes sont en général impossibles, et il est alors difficile de savoir dans quelle mesure le succès d'un système donné est du à la capacité de l'expérimentateur d'ajuster les nombreux paramètres de son système ou à une réelle différence conceptuelle (Tyrrell, 1993b). Il est tentant de proposer une évaluation comparative des systèmes en termes très simples, liés à l'objectif initial de la sélection de l'action, consistant en une mesure de la durée de vie d'un animat doté de divers systèmes de sélection. Le besoin de « bancs d'essai » pour la sélection de l'action est donc très net, le seul proposé (Tyrrell, 1993a) ayant été très peu repris.

Par la mise en place de ces listes de « propriétés souhaitables » trop peu spécifiées, on évite d'aborder de manière directe le problème de l'évaluation du comportement d'ensemble du système de sélection. Cette question n'est en effet pour l'instant pas résolue de façon satisfaisante.

Compte tenu de ces limitations méthodologiques, nous essaierons autant que possible dans nos travaux de mettre en œuvre des comparaisons quantitatives entre les systèmes proposés

et d'autres systèmes plus simples, lors d'ajout de nouvelles fonctionnalités, de comparer les performances des systèmes incluant ou non cet ajout, enfin, lors de la vérification d'une des « propriétés souhaitables », d'adopter une approche quantitative statistique.

1.2 Modèles computationnels de sélection de l'action

De nombreux modèles computationnels de sélection de l'action ont été proposés, tout d'abord dans le cadre de l'éthologie, puis de l'approche animat. Nous passerons brièvement en revue les principales caractéristiques de ces modèles, puis nous nous intéresserons plus particulièrement à leurs éventuelles relations avec des tâches de navigation.

1.2.1 Modèles éthologiques

On peut distinguer deux classes principales de modèles éthologiques (Guillot, 1986) : ceux « de connaissance » et ceux « de représentation ».

Les premiers sont fondés sur des données physiologiques concernant, par exemple, la déplétion énergétique, hydrique et la distension gastrique pour prédire les occurrences ou les enchaînements de comportements alimentaires et dipsiques (Toates, 1986; Guillot, 1988). Cependant, peu de connaissances étant disponibles, ils sont restés d'un nombre très restreint.

Les modèles « de représentation », eux, ne sont pas fondés sur des connaissances physiologiques ou neurologiques précises, mais cherchent paradoxalement à reproduire le comportement animal par des mécanismes spéculatifs influencés par la cybernétique (McFarland, 1971b). Ces mécanismes font référence aux caractéristiques homéostatiques supposées des systèmes motivationnels qui incitent un comportement à s'exprimer pour que l'animal reste dans sa « zone de viabilité » - reprenant en cela les propriétés de l'homéostat d'Ashby (1952).

La plupart de ces modèles reposent sur la notion de « forces des facteurs causaux » (FFC) qui représentent des combinaisons des variables internes et externes conduisant l'animal à exhiber une action ou une autre. Cette notion se rapproche de la notion de motivation telle qu'elle est définie par les éthologues (Toates, 1986) et peut être utilisée en tant que « common currency », qui permettrait de gérer « avec la même monnaie » des comportements disparates (McFarland et Sibly, 1975).

Certains modèles supposent que les FFC correspondant à chaque action sont indépendantes, mais certains autres les rendent interdépendantes –sachant qu'un comportement peut en influencer un autre, comme le rat assoiffé qui diminue ses prises alimentaires (Roper et Crossland, 1982). Dans l'ensemble de ces modèles, c'est toujours l'action avec la FFC la plus grande qui est exprimée, de sorte que des modèles très différents, s'ils ont leurs paramètres bien ajustés, sont capables de reproduire de façon aussi satisfaisante les mêmes séquences comportementales

(Atkinson et Birch, 1970).

Ces modèles computationnels éthologiques, qui ont proliféré jusqu'à la fin des années 1980, n'ont pas validé plus avant leurs résultats, faute de connaissances neurobiologiques sur les mécanismes d'arbitration comportementale.

1.2.2 Modèles ingénieurs

Avec l'essor de l'intelligence artificielle et de l'approche animat, de très nombreux autres mécanismes de sélection de l'action ont été proposés (voir (Tyrrell, 1993a; Pirjanian, 1999; Prescott *et al.*, 1999) pour des revues). Contrairement aux précédents, ils n'ont a priori aucune raison de s'intéresser à une approche biomimétique. Cependant, ils semblent s'être inspirés non pas des modèles computationnels issus de l'éthologie, mais de modèles spéculatifs comme ceux de Lorenz (1950) ou Tinbergen (1951), s'interrogeant notamment sur la notion de centralisation ou de hiérarchie dans l'arbitrage des actions.

Compte tenu des nombreuses approches très différentes qui ont été proposées, il est possible de tirer de leur étude un certain nombre d'enseignements concernant la conception d'un système de sélection de l'action (Pirjanian, 1997; Prescott *et al.*, 1999; Guillot et Meyer, 2000). Dans les paragraphes suivants, seuls des modèles représentatifs sont signalés.

1. La prise de décision peut, tout d'abord, être conçue comme étant décentralisée (e. g. Brooks, 1986 ; Maes, 1991) ou centralisée (e. g. Rosenblatt, 1995). Toute l'approche de la robotique « fondée sur le comportement » (*behavior-based robotics*) dément la nécessité d'un contrôle central de la sélection et propose des modèles inspirés des insectes qui mettent en avant la robustesse accrue d'une sélection répartie lors de la mise hors service d'un composant. Cependant, il semble que l'apparition de mécanismes centralisés de sélection au cours de l'évolution, en particulier chez les vertébrés, soit un avantage lorsque le nombre de comportements à gérer devient important (Prescott, 2001). Par conséquent, le choix entre ces deux approches dépend du degré de complexité que l'on veut pouvoir traiter.

2. Qu'une architecture de sélection de l'action soit centralisée ou non, l'arbitrage a en général lieu entre des modules représentant chacun une action élémentaire. L'utilisation de modules aux fonctions redondantes mais fonctionnant sur des principes différents est recommandée, puisqu'elle permet une plus grande robustesse lorsque l'un des modules est en panne ou dans des conditions où il ne peut fournir une réponse adaptée. Par exemple,

un animat capable d'approcher sa station de recharge soit en se dirigeant vers un signal radio émis par cette station, soit en repérant sa position visuellement, est alors capable de la rejoindre s'il est dans le noir ou si l'émetteur radio est en panne.

3. L'architecture peut être hiérarchique (e.g. Gat, 1991) ou non (e.g. Maes, 1991). On signalera également la proposition, dite de hiérarchie à libre flux (DAMN de Rosenblatt et Payton, 1989), où la structuration de la sélection est hiérarchique mais où les calculs sont pour autant menés en parallèle jusqu'à la décision finale, ce qui permet ainsi d'effectuer une sélection résultant d'un compromis entre plusieurs comportements. Bien que Tyrrell (1993a) ait montré l'avantage d'une hiérarchie à libre flux sur de nombreux autres mécanismes de sélection de l'action, Bryson (2000) a récemment montré la supériorité d'un mécanisme hiérarchique simple sur celui de Tyrrell, dans l'environnement de test de ce dernier. Elle critique en effet les modèles non hiérarchiques en arguant la nécessité d'une structuration hiérarchique pour la gestion de nombreux comportements.

4. Certaines architectures hybrides n'utilisent que des modules réactifs, permettant à un animat de se comporter par réflexe. Or, en cas de ressources rares dans l'environnement, une planification des chemins grâce à une carte mémorisée s'avère nécessaire. Des architectures –appelées hybrides– intègrent donc à la fois des modules réactifs et des modules délibératifs. Elles se trouvent alors face à un problème de coordination de leurs activations, s'interrogeant sur une éventuelle priorité de l'une ou de l'autre. Selon Pirjanian (1997), ce qui apparaît comme plus efficace est de les placer au même niveau dans le processus de sélection. Cela implique de considérer les résultats des calculs de planification comme de simples indications, et non comme des ordres impératifs, et permet alors au système d'adapter ses plans à la dynamique de l'environnement. C'est ce que Agre et Chapman (1990) appellent l'approche « plans-as-resources ». Elle a notamment inspiré la logique employée par l'architecture DAMN de Rosenblatt et Payton (1989). Ce type d'architectures doit être doté de la capacité de fonctionner dans un mode asynchrone afin de pouvoir à la fois mener des calculs complexes de planification tout en réagissant aux modifications soudaines de l'environnement grâce aux modules réactifs.

5. Enfin, la caractéristique de persistance, évoquée précédemment, est probablement l'une des « propriétés souhaitables » les plus importantes. En effet, ces modèles ingénieurs ont

été confrontés de manière chronique au problème des oscillations comportementales (Tyrrell, 1993a).

Il apparaît que les contraintes issues de l'implémentation de ces modèles dans des robots situés a permis la production de mécanismes plus détaillés pour une sélection efficace que les modèles éthologiques qui, eux, ne pouvaient pas être confrontés à un environnement réel ou simulé.

1.2.3 Mécanismes de sélection de l'action et navigation

Les mécanismes de sélection de l'action traitent nécessairement de navigation, dans la mesure où ils s'intéressent à des animats mobiles qu'il faut pouvoir diriger vers des ressources. Cependant, le problème central de la sélection de l'action n'est pas celui de l'implémentation de la navigation en elle-même mais celui du maintien des variables internes dans leur zone de viabilité par le choix des actions adaptées.

C'est pourquoi trois catégories d'architectures se déclarant traiter de sélection de l'action peuvent être identifiées : (1) celles qui se concentrent sur les choix motivationnels et négligent plus ou moins partiellement les problèmes de navigation, (2) celles qui négligent les problèmes de choix motivationnel et intègrent des mécanismes de sélection de l'action pour résoudre exclusivement des problèmes de navigation, (3) celles qui traitent les deux problèmes –navigation et sélection de l'action– simultanément, dans toute leur complexité.

1. Les premières sont représentées dans un grand nombre de modèles (e.g. Maes, 1991). Dans la plupart d'entre eux, les animats ne peuvent rejoindre une ressource qu'à partir du moment où elle entre dans leur champ perceptif. Une simple procédure de guidage ou de remontée de gradient permet en effet dans ce cas de déterminer quelle direction suivre pour s'en approcher. Cette stratégie de navigation, dite d'« approche d'objet », est la plus rudimentaire que l'on puisse concevoir, mais est cependant la seule disponible pour un très grand nombre d'animaux. Elle permet la survie si les ressources sont suffisamment nombreuses et la topologie de l'environnement suffisamment simple pour qu'une exploration aléatoire permette de percevoir une ressource désirée avant que son absence ne s'avère fatale. Elle a l'avantage de pouvoir être très facilement mise en œuvre sur un robot réel. D'autres modèles s'affranchissent du problème de la perception limitée des senseurs et fournissent à tout instant à l'animat les directions permettant de rejoindre les ressources les plus proches (e.g. Seth, 1998). Utilisée en simulation, cette approche est une version simplifiée de la précédente qui occulte les difficultés de la tâche de navigation pour se concentrer uniquement sur la sélection de l'action.

2. La seconde catégorie d'architectures traite la navigation comme un problème de sélection de l'action, l'animat n'ayant pas à choisir entre des buts différents et éventuellement conflictuels générés par divers systèmes motivationnels. C'est le cas, par exemple, de l'architecture DAMN précédemment citée (Rosenblatt et Payton, 1989) qui permet au robot mobile Navlab de choisir entre divers comportements réactifs ou délibératifs d'orientation.

3. A notre connaissance, seuls deux modèles (Gaussier *et al.*, 2000; Guazzelli *et al.*, 1998), et dans une moindre mesure un troisième (Arleo, 2000), traitent les problèmes de navigation et de sélection de l'action avec la même attention. Ils se contraignent à la construction d'une carte cognitive dans des conditions réalistes, à partir des seules données sensorielles accessibles à l'animat. Cette carte peut être exploitée pour rejoindre des ressources hors du champ perceptif, par apprentissage d'une politique en chaque lieu de la carte ou par planification de trajectoire. Ils se contraignent également à l'utilisation d'un « métabolisme virtuel » limitant le temps disponible pour la cartographie par les impératifs de survie. Nous les décrirons de façon plus détaillée au chapitre 4.

1.2.4 Bilan

Ce bref survol des modèles éthologiques et ingénieurs permet de dégager les points essentiels qui ont été à l'origine de notre travail.

D'une part, il apparaît que les nombreux mécanismes mis en œuvre dans ces modèles, aussi différents soient-ils, peuvent pour la plupart assurer une sélection de l'action efficace. Très peu de travaux en font une comparaison –à part Tyrrell (1993b), pour des animats simulés. Dans le cadre de l'approche animat, on peut alors s'interroger sur la similarité de certains de ces mécanismes avec ceux mis en place au cours de l'évolution des espèces, qui se sont révélés robustes et efficaces. Recenser les modèles computationnels relevant de cette catégorie et tester les hypothèses neurobiologiques qui sous-tendent les mécanismes de sélection de l'action dans le cadre d'une tâche de survie fera l'objet de la première partie de notre travail.

D'autre part, il apparaît que les architectures de contrôle traitant sur le même niveau navigation et sélection de l'action sont rares, alors qu'un animal, en particulier, et un système autonome, en général doivent gérer ces problèmes de manière conjointe. Alors que, dans la première partie

de notre travail, la navigation a été réduite à une exploration aléatoire de l'environnement, la deuxième partie de notre travail consistera à tester les hypothèses neurobiologiques qui sous-tendent une interface entre plusieurs stratégies de navigation et un système de sélection de l'action.

Chapitre 2

Les ganglions de la base : biologie et modèles computationnels

The best material model of a cat is another, or preferably the same, cat
Rosenblueth & Wiener (1945)

Depuis le milieu des années 90, une nouvelle catégorie de modèles de sélection de l'action ont pu être élaborés. Ils ont à la fois un lien de parenté avec les modèles éthologiques « de connaissance » –de par l'usage qu'ils font des nouvelles données neurobiologiques disponibles concernant les mécanismes de sélection de l'action chez les animaux– et avec les modèles ingénieurs –de par l'ajout de mécanismes non-biomimétiques lorsque les connaissances ne sont pas suffisantes ainsi que par leur évaluation dans des expériences situées. Une attention toute particulière a été portée aux ganglions de la base, un ensemble de noyaux sous-corticaux semblant avoir une fonction générique de sélection.

Nous décrirons tout d'abord l'essentiel de ces structures biologiques, permettant de comprendre les mécanismes sous-jacents. Nous passerons ensuite en revue les modèles computationnels qui en sont inspirés et les différentes fonctions de sélection qui ont été modélisées.

2.1 Données neurobiologiques

Les ganglions de la base (GB) sont un ensemble de noyaux subcorticaux interconnectés, éléments d'une boucle cortex-ganglions de la base-thalamus-cortex (fig. 2.1). Ils sont également intimement liés au système dopaminergique mésencéphalique. Un grand nombre de données expérimentales a été accumulé concernant la structure des ganglions de la base et la physiologie des neurones les composant (voir (Parent et Hazrati, 1995a; Parent et Hazrati, 1995b; Mink, 1996; Wickens, 1997; Houk *et al.*, 1995b; Greenberg, 2001) pour des revues). Cet intérêt particulier est lié au fait que de nombreux désordres du mouvement (maladie de Parkinson,

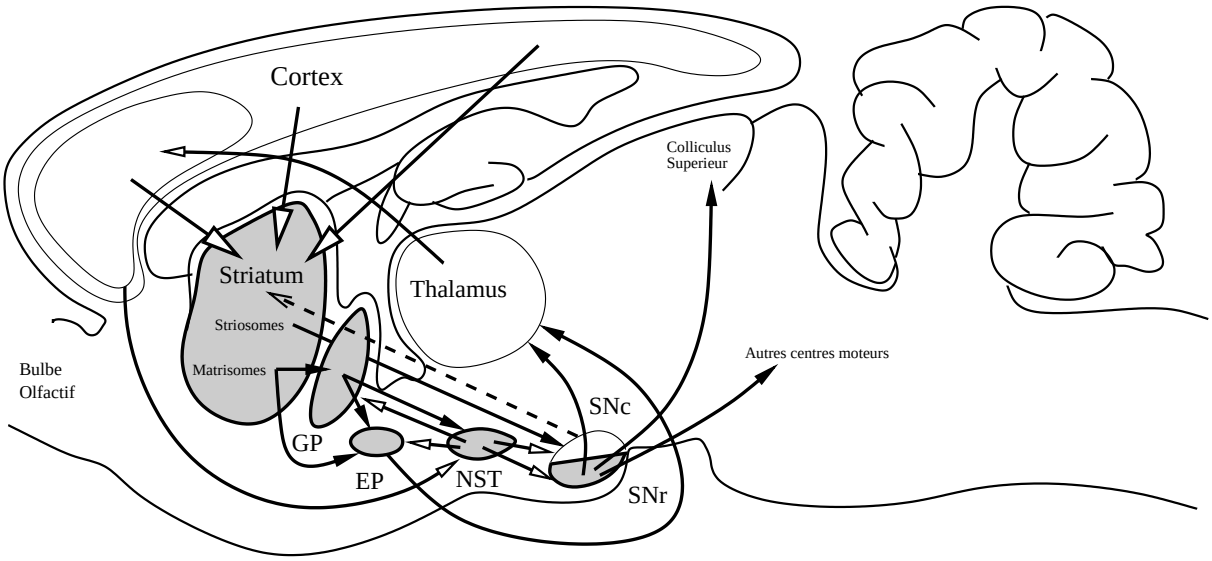

FIG. *2.1: Vue d'ensemble des noyaux de la partie dorsale des ganglions de la base (en grisé), de leurs interconnexions et de leurs connexions avec les noyaux dopaminergiques, chez le rat. Flèches évidées : connexions excitatrices ; flèches pleines : connexions inhibitrices ; flèche en pointillés : connexion dopaminergique.*

maladie de Huntington, syndrome de Tourette, etc.) chez l'humain sont dus à des maladies affectant les ganglions de la base et le système dopaminergique associé.

Cette partie se concentre sur les données principales permettant d'appréhender le fonctionnement des ganglions de la base et d'introduire leurs modèles computationnels.

On peut, dans un premier temps, distinguer deux circuits relativement isolés des ganglions de la base, l'un regroupant des noyaux placés en position dorsale et l'autre en position ventrale.

2.1.1 Circuit dorsal

Anatomie

Les noyaux constituant les circuits dorsaux des GB chez le rat sont organisés comme suit (voir fig. 2.2, gauche) :

- *entrées* : striatum dorsal (ou néostriatum) et partie latérale du noyau subthalamique (NST),
- *noyaux intermédiaire* : globus pallidus (GP),
- *sorties* : noyau entopédonculaire (EP) et la partie latérale de substance noire reticulée (SNr).

Le striatum dorsal est le lieu de convergence de nombreuses excitations (connexions glutamatergiques) issues de l'ensemble du cortex (McGeorge et Faull, 1989). Il est composé à 97%

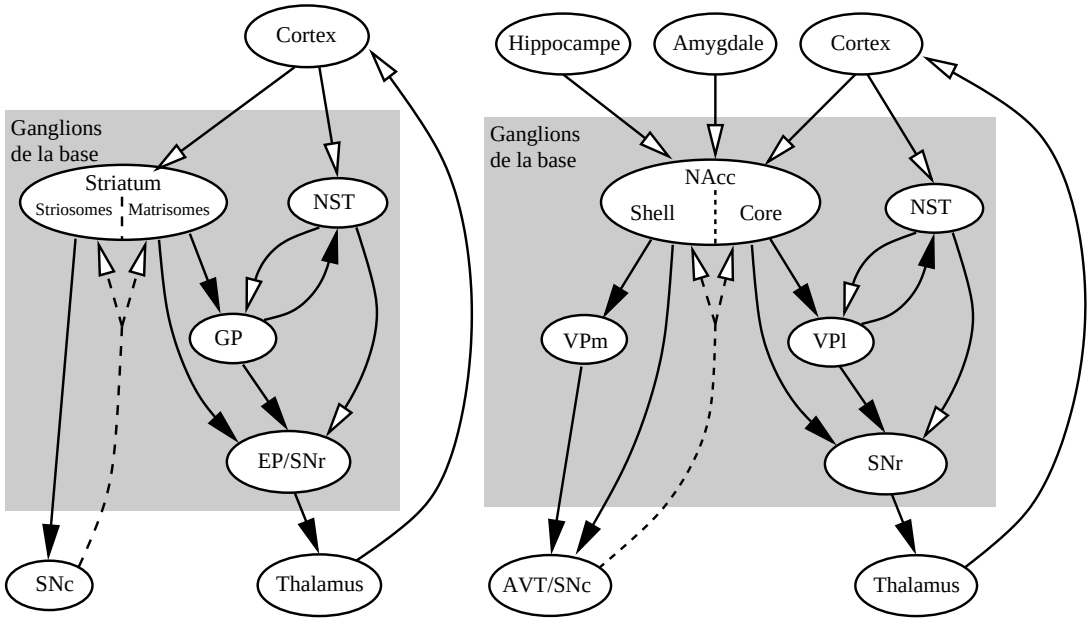

FIG. *2.2: Représentation simplifiée des principales connexions des circuits dorsaux (gauche) et ventraux (droite) des ganglions de la base. Les interconnexions entre les matrisomes du NAcc « core » et la SNc ne sont pas représentées. Les connexions excitatrices sont représentées par des flèches évidées, les connexions inhibitrices par des flèches pleines, les connexions dopaminergiques en pointillés. Les abréviations sont explicitées dans le texte et dans l'annexe A.*

de neurones épineux moyens (MSN), GABAergiques (inhibiteurs). Ces MSN sont en général considérés comme divisés en deux groupes distincts, selon qu'ils possèdent des récepteurs à dopamine de type D1 ou D2.

Il est constitué d'une « matrice » subdivisée en compartiments neuronaux isolés, les matrisomes, au sein de laquelle se distinguent d'autres compartiments aux efférences distinctes, les striosomes.

L'existence d'inhibitions latérales entre matrisomes est sujette à controverse (Jaeger *et al.*, 1994). Ils se projettent sur le globus pallidus, le noyau entopédonculaire et la substance noire réticulée par des connexions GABAergiques, inhibitrices. Le globus pallidus inhibe (GABA) le noyau subthalamique, le noyau entopédoncumaire et la substance noire réticulée. Le noyau subthalamique reçoit également des projections corticales glutamatergiques et excite (glutamate également) le globus pallidus, le noyau entopédonculaire et la substance noire réticulée (voir fig. 2.2, gauche). L'ensemble de ces interconnections conservent l'organisation en compartiments issue des matrisomes, sauf les excitations issues du noyau subthalamique qui s'exercent de façon diffuse sur le GP, l'EP et la SNr. On peut donc identifier dans les ganglions de la base une structuration en canaux distincts (fig.2.3, gauche). Ces interconnexions des noyaux des GB, ont

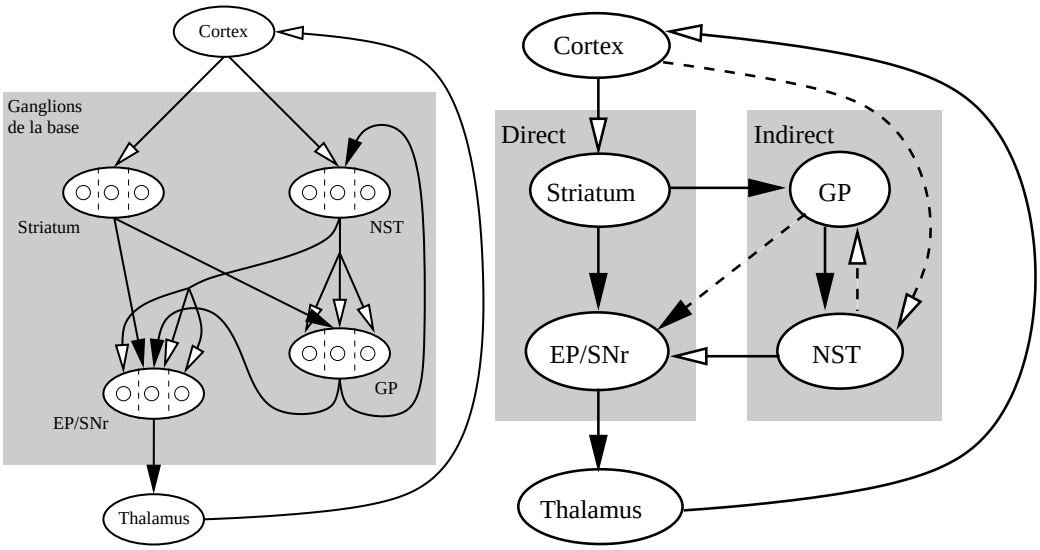

FIG. *2.3: Gauche : Structuration en canaux parallèles des ganglions de la base, seules les projections issues du NST ne respectent pas la ségrégation des canaux (trois canaux sont représentés, seules les projections issues du canal central sont représentées). Droite : Interprétation des connexions des circuits dorsaux des ganglions de la base en terme de chemin direct et chemin indirect. Les connexions excitatrices sont représentées par des flèches évidées, les connexions inhibitrices par des flèches pleines, les connexions en pointillés sont négligées par l'interprétation.*

longtemps été interprétées comme la divergence depuis le striatum puis la convergence dans le complexe EP/SNr d'un chemin direct inhibiteur et d'un chemin indirect excitateur (Albin *et al.*, 1989) (fig. 2.3, droite).

Les striosomes se projettent sur le système dopaminergique mésencéphalique. Ce système est composé principalement de la substance noire compacte (SNc) et de l'aire tegmentale ventrale (AVT). En ce qui concerne la partie dorsale des GB, les striosomes se projettent exclusivement sur la SNc, qui se projette en retour sur l'ensemble du striatum (matrisomes et striosomes, voir fig. 2.2, gauche).

Enfin, les noyaux de sortie (EP/SNr) projettent leurs sorties vers certains systèmes moteurs non-corticaux (colliculus supérieur, etc.), vers divers noyaux du tronc cérébral (noyau pédonculopontin tégmental, etc.) et principalement vers le lobe frontal du cortex via divers noyaux thalamiques (complexe thalamique ventroantérieur-ventrolatéral, VA/VL, et noyau ventromédian, VM). Les zones corticales cibles des circuits dorsaux chez le rat sont :

– le cortex latéral agranulaire (AGl) équivalent du cortex moteur primaire (M1) des primates,
– le cortex médian agranulaire (AGm) équivalent de l'aire motrice supplémentaire, du cor-

tex prémoteur et des champs oculaires frontaux (SMA, PMC et FEF, respectivement) des primates,
– l'aire cingulaire antérieure (ACA) équivalent du cortex préfrontal associatif des primates,

Neurophysiologie

Les neurones de l'EP et de la SNr sont toniques et inhibiteurs, ils maintiennent donc le thalamus et les noyaux du tronc cérébral cibles sous une inhibition gobale. Sous l'effet de stimulations corticales, l'activation de neurones du striatum tend à inhiber sélectivement certains canaux de l'EP et de la SNr, relâchant ainsi de façon ciblée cette constante inhibition. Le fonctionnement des GB est donc compris comme une « désinhibition sélective » de certains canaux (Chevalier et Deniau, 1990), permettant ainsi au thalamus d'exciter des zones circonscrites du cortex.

2.1.2 Circuit Ventral

Les noyaux constituant les circuits ventraux des ganglions de la base sont (voir fig. 2.2, droite) :
– *entrées* : striatum ventral (comprenant le noyau accumbens et la partie profonde du bulbe olfactif) et partie médiale du noyau subthalamique (NST),
– *noyaux intermédiaires* : le pallidum ventral (VP), lui-même divisé en deux parties distinctes, l'une médiale et ventrale (abrégée VPm pour des raisons historiques), l'autre dorsale et latérale (VPl),
– *sorties* : partie médiale de la substance noire réticulée (SNr).

Le noyau d'entrée du circuit ventral des ganglions de la base est le noyau accumbens (NAcc). Ce noyau n'est pas homogène et apparaît divisé en deux parties principales, le noyau ou « core » (en position médio-dorsale) et l'écorce ou « shell » (ventro-latérale) (Zahm et Brog, 1992). Nous considérerons cette dichotomie, malgré la présence d'une zone de transition entre ces deux éléments, le pôle rostral (Heimer *et al.*, 1997), ainsi que de résultats tendant à mettre à jour des subdivisions plus fines au sein de ces divisions (Ikemoto, 2002). Le « core » est considéré comme une simple extension ventrale du striatum dorsal, alors que le « shell » est considéré comme une zone de transition entre le striatum et l'amygdale (Groenewegen *et al.*, 1996; Heimer *et al.*, 1997).

En plus d'afférences standard (cortex préfrontal, thalamus, aire ventrale tégmentale –noyau dopaminergique– et noyau raphé médian –noyau sérotoninergique), les circuits ventraux en possèdent de très spécifiques, issues du système limbique : la formation hippocampique, l'amyg-

dale, le cortex parahippocampique et le cortex antérieur cingulaire (Groenewegen *et al.*, 1999; Groenewegen *et al.*, 1996). Cette particularité fait que ces circuits ventraux sont également nommés circuits limbiques. On notera que les projections issues de l'hippocampe ciblent principalement la partie « shell » du noyau accumbens, alors que celles issues d'aires limbiques du cortex ciblent principalement le « core » (Thierry *et al.*, 2000).

Ces deux régions du noyau accumbens sont en amont de deux circuits distincts des ganglions de la base (Thierry *et al.*, 2000) (voir fig. 2.2, droite). Les noyaux des ganglions de la base situés en aval du « core » du noyau accumbens ont une anatomie et une électrophysiologie tout à fait semblables à celles des noyaux situés en aval du striatum dorsal.

En effet, d'un point de vue anatomique, la partie dorso-médiale de la substance noire réticulée (SNr) semble jouer, pour ce circuit, le rôle de sortie joué par le complexe EP/SNr pour les circuits dorsaux des ganglions de la base (Maurice *et al.*, 1999). Le pallidum ventral dorsal et latéral (VPl) y joue un rôle équivalent au globus pallidus (GP) des circuits dorsaux (Maurice *et al.*, 1997). La région médiale du noyau sub-thalamique (NST) est dédiée à ce circuit ventral, fournissant des excitations diffuses vers le NAcc « core » et le VPl (Parent et Hazrati, 1995b) (voir fig. 2.2). Enfin, bien qu'il reste encore à en établir la preuve formelle, il est supposé que les neurones du NAcc « core », à l'instar de ceux du striatum dorsal, sont organisés en trois sous-populations, l'une se projetant exclusivement sur la SNr, l'autre sur le VPl et la dernière vers les deux noyaux, ainsi que sont organisés les projections des neurones du striatum dorsal vers EP/SNr et GP (Maurice *et al.*, 1997).

Ce circuit ventral issu du « core » cible principalement le noyau médiodorsal du thalamus (MD), qui se projette sur l'aire agranulaire insulaire (AIA), en général assimilée aux aires préfrontales limbiques (orbitofrontale, infralimbique et prélimbique) des primates.

En ce qui concerne l'électrophysiologie de ce circuit, des stimulations corticales (cortex prélimbique et médial orbital, PL/MO) produisent des patrons de réponse similaires à ceux obtenus suite à des stimulations du cortex sensorimoteur (Maurice *et al.*, 1999).

2.1.3 Rôle des ganglions de la base

Le rôle précis des ganglions de la base dans le système nerveux central est sujet à controverse. Ils ont initialement été considérés comme faisant partie du système moteur extrapyramidal, c'est-à-dire de la partie du système moteur en charge des aspects automatiques du mouvement (Mink, 1996). Les études électrophysiologiques et celles des lésions des GB confirment qu'ils ont un rôle dans les fonctions motrices de bas niveau, ceci malgré une absence de connexion directe avec la moelle épinière. Par ailleurs, de nombreuses pathologies humaines produisant des troubles du contrôle moteur sont liées à des désordres dans le fonctionnement de GB : maladie de Parkinson, maladie de Huntington, syndrome de Tourette, ballisme, dyskinesie

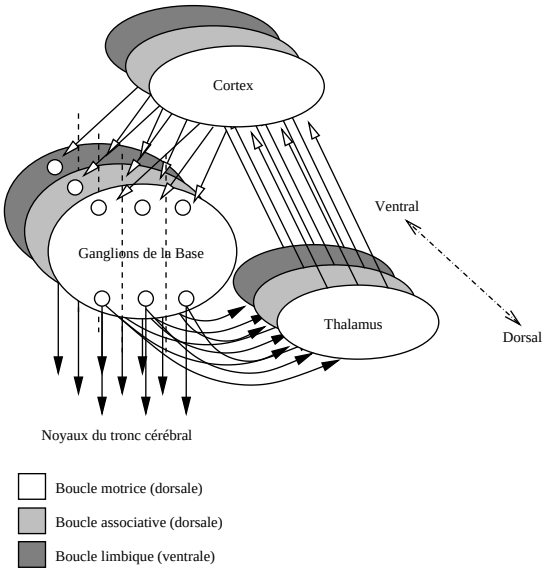

FIG. *2.4: Représentation schématique des trois grandes boucles cortex-ganglions de la base-thalamus-cortex. Pour chacune des boucles, trois canaux sont représentés*

tardive. Pour autant, le niveau auquel a lieu cette sélection (comportement intégré, primitive comportementale, etc.) n'est pas clairement défini.

Cependant, de nombreux travaux suggèrent également que les GB ont un rôle de sélection étendu à des fonctions de plus haut niveau (Middleton et Strick, 1994; Greenberg, 2001) : sélection des comportements, motivation, planification, apprentissage de la récompense associée à une action, mémoire à court terme, catégorisation. Là aussi, un certain nombre de maladies neuropsychiatriques sont associées aux dysfonctionnements des GB (Wickens, 1997) : schizophrénie, troubles compulsifs obsessionnels, hyperactivité et déficit attentionnel, divers déficits de l'apprentissage, etc.

La fonction des ganglions de la base pourrait s'apparenter à la sélection, parmi un grand nombre d'activations du cortex, de celles qui sont pertinentes en fonction du contexte. Les GB sont alors considérés comme un interrupteur de type « qui perd gagne » : le canal d'entrée soumis à l'excitation corticale la plus forte est désinhibé, ce qui permet un renforcement de cette activation corticale maximale.

2.1.4 Boucles parallèles

Les boucles cortex-ganglions de la base-thalamus-cortex, dorsale et ventrale, sont subdivisées en sous-boucles indépendantes, chacune issue de régions différentes du cortex, passant par des sous-parties spécifiques de noyaux des ganglions de la base et du thalamus (Alexander *et al.*, 1986; Alexander *et al.*, 1990; Alexander et Crutcher, 1990) (voir fig. 2.4). Elles sont

disposées sur un axe ventro-dorsal et sont fonctionnellement différenciées, ce qui explique l'implication des ganglions de la base à la fois dans des processus moteurs et cognitifs. Chez le rat, on distingue trois boucles principales indépendantes, au sein desquelles d'autres subdivisions sont susceptibles d'exister :

1. La boucle motrice : il s'agit de la sous-partie dorsale des circuits dorsaux des ganglions de la base décrits plus haut. Impliquant des régions motrices du cortex, elle aurait un rôle dans la mémoire procédurale et l'apprentissage de comportements « habitudes », de type S-R (Graybiel, 1998; Cardinal, 2001; Everitt et Wolf, 2002).
2. La boucle associative : il s'agit de la sous-partie ventrale des circuits dorsaux. Elle aurait un rôle dans la mémoire de travail et l'apprentissage de séquences.
3. La boucle limbique : elle correspond au circuit ventral des ganglions de la base décrit précédemment. Issue du cortex limbique, c'est la seule qui possède des afférences de l'hippocampe et de l'amygdale. Elle se subdivise en deux sous-circuits, l'un issu du « shell » du noyau accumbens, l'autre du « core ». Celui issu du « shell » prendrait part aux processus de motivation et de récompense (Kelley, 1999). Celui issu du « core » serait plutôt dédié à la génération de réponses comportementales en situation nouvelle, lorsque les « habitudes » sélectionnées par les boucles dorsales ne sont pas adaptées (Ikemoto et Panksepp, 1999). En particulier, dans le cadre de la navigation, lorsqu'un trajet habituel n'est plus valide suite à une modification de l'environnement (déplacement de la récompense dans un labyrinthe, par exemple), l'utilisation de processus de navigation et de planification de trajectoire pour générer de nouvelles actions locomotrices semble prendre place dans ce circuit (Seamans et Phillips, 1994).

La mise en œuvre d'un système de sélection de l'action inspiré des ganglions de la base semble donc devoir impliquer prioritairement des circuits moteurs. L'intégration de capacités de navigation dans ce même modèle nécessitera son extension à la sous-boucle de la boucle limbique ayant pour entrée la partie « core » du noyau accumbens (que l'on dénommera boucle limbique « core »).

2.1.5 Communication inter-boucles

L'existence de boucles « parallèles » pose le problème de leurs interconnexions, ces dernières apparaîssent en effet nécessaires pour coordonner leurs actions. En particulier, le rôle joué par la boucle limbique « core » dans la prise en compte de la navigation implique qu'elle soit capable d'influer directement ou indirectement sur la locomotion. Elle est donc susceptible d'être en conflit avec des ordres de locomotion, ou au contraire d'arrêt, issus de la boucle motrice. Plusieurs hypothèses concernant les interconnexions entre boucles ont été proposées :

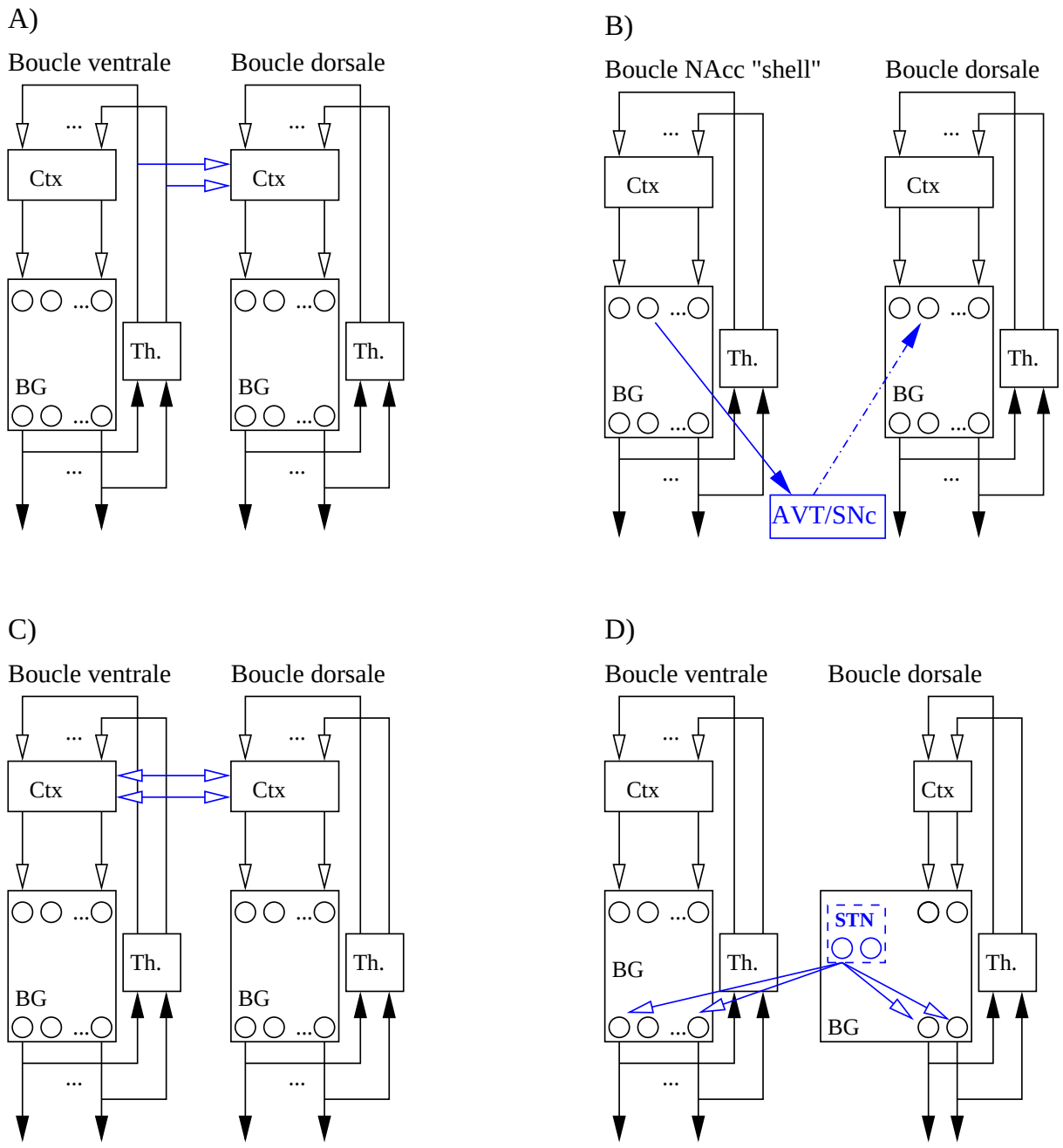

FIG. 2.5: *Interconnexions éventuelles entre boucles des ganglions de la base. A) voie hiérarchique, B) voie hiérarchique dopaminergique, C) voie cortico-corticale, D) voie trans-subthalamique. Flèches pleines : connexions inhibitrices ; flèches évidées : connexions excitatrices ; flèches en pointillés : connexions dopaminergiques.*

– *La voie hiérarchique* (Joel et Weiner, 1994) : cette hypothèse suggère que les noyaux thalamiques d'une boucle se projettent non-seulement sur les zones du cortex participant de cette boucle, mais également sur celles de la boucle située au niveau inférieur dans la hiérarchie (fig. 2.5, A). Les boucles limbiques sont supposées être au plus haut niveau de la hiérarchie, suivies par les circuits associatifs puis moteurs. L'existence de ces connexions au niveau thalamo-cortical est sujette à controverse, en particulier en ce qui concerne l'influence des circuits limbiques sur les autres. Enfin, ces interconnexions ne permettent la transmission d'information que dans le sens ventral vers dorsal.

– *La voie hiérarchique dopaminergique* (Joel et Weiner, 2000) : chez le rat, les boucles motrices et associatives sont en interaction avec des neurones dopaminergiques, respectivement ceux de la substance noire compacte latérale et de la substance noire compacte médiale. Ces interactions consistent en des projections des striosomes de la sous-partie du striatum de la boucle considérée sur ces neurones dopaminergiques, qui en retour se projettent sur l'ensemble de cette sous-partie du striatum –striosome et matrisomes. Via les striosomes du NAcc « core », la boucle limbique correspondante suit un schéma similaire de connexion réciproque avec l'aire ventrale tégmentale (AVT). Cependant, la boucle limbique « shell », elle, est en mesure de moduler l'activité des boucles motrices, associatives et limbique « core » via des projections vers les zones ventrales de la SNc et vers l'AVT (fig. 2.5, B). Encore une fois, cette voie ne permet la transmission d'information que dans le sens ventral vers dorsal.

– *La voie cortico-corticale* : L'éventualité d'interconnexions au niveau cortical entre des aires en interaction avec des boucles distinctes ouvre la voie à un transfert bilatéral d'information (fig. 2.5, C). Cependant, chez le rat, ces interconnexions sont relativement limitées, en particulier les interconnexions entre le cortex prélimbique et les aires postérieures (motrices ou sensorielles entre autres) (Preuss, 1995).

– *La voie trans-subthalamique* (Kolomiets *et al.*, 2001; Kolomiets *et al.*, 2003) : La ségrégation entre boucles semble n'être pas totalement conservée au niveau du noyau subthalamique : une partie de ses neurones ont des afférents d'aires corticales appartenant à des boucles différentes, de sorte qu'une région de la substance noire réticulée associée à une

boucle est susceptible d'être excitée par une autre boucle (fig. 2.5, D). Ces excitations viennent renforcer l'inhibition de sortie exercée par la SNr, tendant à bloquer la sélection dans la boucle subissant cette excitation supplémentaire. Cette capacité d'une boucle à éteindre une partie de la sélection d'une autre boucle est un processus de coordination entre boucles qui n'est pas soumis à une structuration hiérarchique, mais qui reste cependant limité à de petites régions de la SNr, à la frontière entre deux zones dédiées à des boucles différentes.

– Enfin, on ne peut écarter la possibilité que les conflits entre boucles ne soient pas réglés dans les boucles des GB, mais dans les structures neurales situées en aval, ce qui rendrait caduque la question des interconnexions entre boucles.

2.1.6 Rôle de la Dopamine

Signal de renforcement

Les études électrophysiologiques du système dopaminergique menées par Schultz (Schultz, 1986; Romo et Schultz, 1990) ont révélé un lien fort entre ce système et l'apprentissage par essai-erreur (ou apprentissage par renforcement, voir 2.2.1 pour une définition plus précise). En effet, les neurones dopaminergiques déchargent en réponse à des récompenses inattendues. De plus, lors de tâches d'apprentissage comportemental, ils ne déchargent à l'arrivée de la récompense (stimulus inconditionnel) qu'au début des expériences et décalent, au fur et à mesure de l'acquisition de la tâche, leur activation vers le moment où arrive le stimulus prédisant le renforcement (stimulus conditionnel). Enfin, la dépression et la potentiation à long terme des réponses synaptiques dans le striatum ont été observées (Wickens, 1997), cette plasticité synaptique pouvant être la base d'un mécanisme d'apprentissage basé sur la dopamine. La dopamine est donc couramment considérée comme le *médiateur du renforcement* (Wickens et Kötter, 1995). Doya estime d'ailleurs que la résolution des problèmes d'apprentissage par renforcement est la caractéristique principale et spécifique des ganglions de la base, qui distingue clairement leur rôle de ceux du cortex et du cervelet (Doya, 1999).

Signal incitateur, signal de nouveauté

Cette vision est cependant critiquée, en particulier par Berridge et Robinson (1998), qui observent entre autres que les protocoles expérimentaux utilisés dans ces tâches d'apprentissage nécessitent tous de l'animal une phase d'approche (se tourner vers un stimulus visuel, s'avancer vers un levier, etc.). Ils ont développé l'idée que la dopamine est en réalité un signal qu'ils nomment *salience incitatrice.* Cette *salience incitatrice* serait utilisée par une sous-partie de la

tâche globale d'apprentissage, qui aurait pour rôle d'attribuer à des objets de l'environnement un caractère attractif.

Enfin, la distinction entre les sources de dopamine –substance noire compacte et aire ventrale tégmentale (SNc et AVT)– et leurs afférents (partie ventrale ou dorsale des ganglions de la base) ne doit pas être négligée. En effet, l'activation de la partie du système dopaminergique issu de la SNc génère des comportements stéréotypés (flairage, toilette, mouvements de la mâchoire), alors que l'activation sélective du système dopaminergique issue de l'AVT augmente l'activité locomotrice mais pas la génération de ces comportements stéréotypés (Honkanen, 1999) et joue un rôle fondamental dans la modulation du comportement par les motivations (en particulier la faim) (Kelley, 1999). Ces différences concernant les effets induits par la dopamine issue de la SNc et celle issue de l'AVT invitent à considérer que la dopamine pourrait avoir des rôles divers. Ainsi, Ikemoto et Panksepp (1999) considèrent la dopamine de l'AVT comme médiatrice d'un *signal de nouveauté* permettant de distinguer les moments où les comportements habitudes sont adaptés et ceux où il faut inciter la génération de comportements exploratoires de recherche de récompense, développant et élargissant ainsi la proposition de Berridge et Robinson.

Signal de transition comportementale

Enfin, Redgrave *et al.* (1999b) considèrent que la dopamine contrôle la dynamique de la sélection effectuée par les circuits des ganglions de la base. Un taux de dopamine inférieur à la normale bloquerait la capacité à effectuer une sélection, alors qu'un taux élevé faciliterait la sélection au point que plusieurs canaux pourraient être sélectionnés en simultané.

Le rôle de la dopamine apparaît donc comme complexe et actuellement non-élucidé. De plus, les hypothèses présentées ne sont pas toutes exclusives, en particulier si l'on envisage la possibilité que les modulations dopaminergiques n'aient pas le même rôle selon leur origine (SNc ou AVT), selon les boucles cortex-ganglions de la base-thalamus-cortex qu'elles ciblent et enfin selon que l'on considère l'activité phasique ou tonique des neurones dopaminergiques.

2.2 Modèles computationnels des GB

De nombreux modèles computationnels des ganglions de la base ont été proposés depuis le début des années 90 (voir (Houk *et al.*, 1995b; Wickens et Kötter, 1995; Beiser *et al.*, 1997; Gillies et Arbruthnott, 2000; Joel *et al.*, 2002) pour des revues). Ces modèles couvrent l'ensemble des fonctions des boucles cortex-ganglions de la base-thalamus-cortex.

Une grande part d'entre eux s'intéresse essentiellement aux fonctions d'apprentissage des ganglions de la base en relation avec les noyaux dopaminergiques, du fait du rôle supposé de la dopamine dans l'apprentissage (voir 2.1.6). Notre travail n'a pas pour objectif de doter Psikharpax de capacités d'apprentissage. Cependant, ces mécanismes d'apprentissage étant intégrés à de nombreux autres modèles traitant plus spécifiquement des rôles des ganglions de la base, nous commencerons par les présenter brièvement.

Les autres modèles seront ensuite classés selon les différentes fonctions des GB auxquelles ils s'intéressent : la mémorisation à court terme, la génération de séquences et le contrôle de trajectoire bas niveau.

L'ensemble de ces modèles présente un intérêt dans la modélisation de la sélection de l'action. La capacité à conserver des informations contextuelles en mémoire à court terme est utilisée par de nombreux systèmes de sélection de l'action, par exemple pour contrôler le déroulement des étapes d'un comportement élémentaire. La capacité à générer des séquences est l'une des « propriétés souhaitables » (cf. 1.1.4) attendues d'un mécanisme de sélection de l'action. Enfin, le contrôle de trajectoire est situé en aval des mécanismes de sélection de l'action. La réalisation d'un système biomimétique tel que Psikharpax sera donc à terme confronté à ces problématiques. Enfin, l'intégration de capacités d'apprentissage par renforcement dans les mécanismes de sélection de l'action est une étape qui devra nécessairement succéder à la construction de systèmes ajustés à la main.

2.2.1 Apprentissage par renforcement

Définition

L'apprentissage par renforcement, dans l'acception issue du domaine de l'apprentissage machine, est le problème auquel est confronté un agent qui doit apprendre son comportement par des interactions de type essai-erreur avec un environnement dynamique (Sutton et Barto, 1998; Kaelbling *et al.*, 1996). Il considère un animat en interaction avec son environnement via ses perceptions et les actions qu'il peut entreprendre. Ses perceptions sont de deux types : celles qui renseignent sur l'état de l'environnement, ainsi qu'un signal de renforcement unique, assimilable à une récompense ou une punition, informant l'agent de la qualité de son comportement global. L'objectif de cet apprentissage est de doter l'agent d'un comportement maximisant la somme à long terme des signaux de renforcement. L'apprentissage par renforcement se distingue donc de l'apprentissage supervisé classique, qui consiste en la présentation de couples « perception/action correcte » à un système chargé d'apprendre à les reproduire. Cet apprentissage a en général lieu « hors ligne », avant l'utilisation et l'évaluation du système. L'apprentis-

sage par renforcement suppose que l'agent apprend « en ligne », c'est-à-dire tout en fonctionnant réellement dans son environnement, et qu'il ne dispose que d'une information très pauvre sur la qualité de son comportement, le signal de renforcement, qui ne lui indique en rien les actions correctes à effectuer.

Modèle acteur/critique

La dopamine étant considérée par de nombreux auteurs comme le neuromédiateur du renforcement (voir 2.1.6), toute une branche de la modélisation des GB s'intéresse à modéliser le système biologique en charge de l'apprentissage par renforcement. La forte ressemblance entre l'activité des neurones dopaminergiques et le signal d'erreur de prédiction du modèle d'apprentissage par différence temporelle (« TD Learning ») de Sutton et Barto (1998) a donné lieu à des modèles assimilant le niveau de dopamine à ce signal d'erreur (Houk *et al.*, 1995a; Montague *et al.*, 1996). Le modèle d'apprentissage par différence temporelle est basé sur une structure duale acteur/critique, où l'acteur contrôle les actions du système et où le critique reçoit le renforcement en provenance de l'environnement et prédit la somme des récompenses futures. Les erreurs de prédiction du critique sont utilisées, d'une part, par l'acteur qui va augmenter ou diminuer sa propension à effectuer l'action qui a généré l'erreur (selon que la récompense aura été plus ou moins importante que la prédiction), et d'autre part, par le critique, qui va modifier sa prédiction dans le contexte qui a généré l'erreur de sorte à ne pas la reproduire.

Les modèles biomimétiques d'apprentissage par renforcement proposent d'identifier les striosomes et les noyaux dopaminergiques au critique, les variations de décharge de dopamine à l'envoi de signaux d'erreur de prédiction, et le reste des ganglions de la base à l'acteur (Barto, 1995; Houk *et al.*, 1995a; Schultz *et al.*, 1997; Salum *et al.*, 1999). Divers raffinements et améliorations ont depuis été ajoutés à ce modèle de base afin qu'il puisse rendre compte d'un maximum de faits expérimentaux (Brown *et al.*, 1999; Contreras-Vidal et Schultz, 1999; Suri et Schultz, 1999; Bar-Gad *et al.*, 2000; Bar-Gad et Bergman, 2001; Suri et Schultz, 2001; Baldassarre, 2001). Une revue récente de l'ensemble de ces modèles est disponible dans (Joel *et al.*, 2002).

On notera que la modélisation de la partie acteur est en général limitée à sa plus simple expression (un perceptron où chaque neurone artificiel code pour une action chez (Houk *et al.*, 1995a)), malgré la complexité des circuits correspondants (voir 2.1.1 et 2.1.2). Seul le modèle de Berns et Sejnowski (1996) fait exception de par sa partie acteur développée. Chaque canal des ganglions de la base y est assimilé à une action élémentaire, la ségrégation en canaux est présente dans l'ensemble des noyaux à l'exception du NST, représenté par un unique neurone (voir fig. 2.6, gauche). Dans ce modèle, fondé sur l'interprétation chemin direct/chemin indirect (fig. 2.3, droite), le chemin direct sert à inhiber dans le globus pallidus interne les actions en

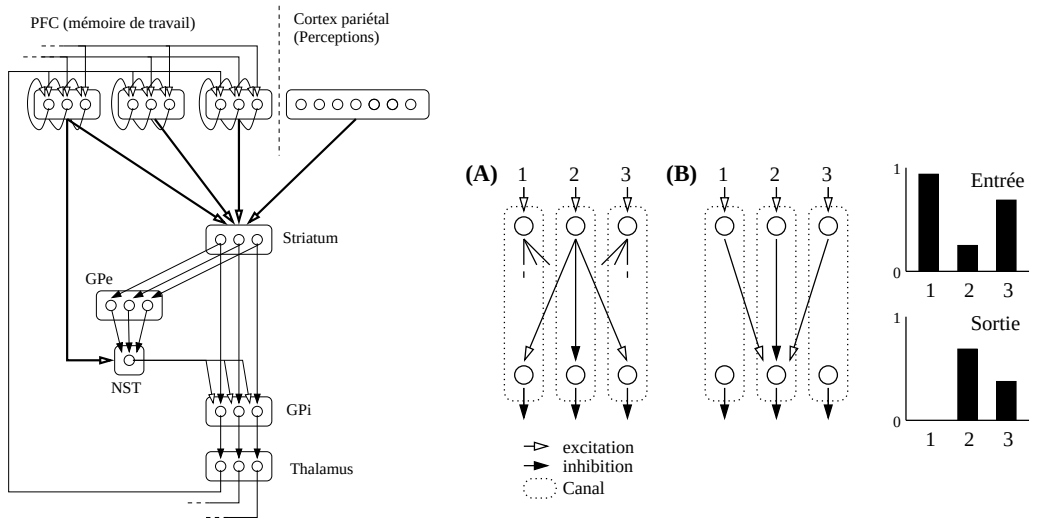

FIG. *2.6: Gauche : schéma de la partie acteur du modèle de Berns & Sejnowski ; en gras : connexions n :m ; flèches évidées : excitations ; flèches pleines : inhibitions. Droite : principe de fonctionnement d'un réseau « off-center on-surround ». (A) Projection du neurone d'entrée d'un canal ; (B) Afférences du neurone de sortie d'un canal.*

fonction de leur niveau d'excitation en entrée, alors que le chemin indirect génère une excitation globale de toutes les actions dans le globus pallidus interne. De la combinaison des deux résulte la sélection par désinhibition d'une seule action. Ce type d'architecture est semblable aux réseaux connexionnistes « on-center off-surround » utilisés pour leur capacité de sélection « qui perd gagne » : pour chaque canal, le neurone d'entrée inhibe le neurone de sortie qui lui correspond et excite les autres. La résultante de cette architecture est que le canal le plus excité en entrée excite fortement ses voisins, tout en étant capable d'inhiber totalement les excitations qu'il reçoit (voir fig. 2.6, droite). Ce modèle a été testé sur des tâches cognitives simplifiées, dans des configurations *ad hoc* (WSCT et de Bandit multi-bras, voir annexe B).

Malgré ces nombreux travaux dérivés des algorithmes d'apprentissage TD, l'assimilation du signal dopaminergique au signal d'erreur de prédiction ne fait pas l'unanimité. Nous signalerons donc que Pennartz a identifié de nombreux points sur lesquels ces algorithmes ne sont pas biologiquement plausibles (Pennartz, 1996; Pennartz *et al.*, 2000). Il propose par conséquent un modèle d'apprentissage par renforcement (Pennartz, 1997) indépendant de la dopamine, nommé HSAT pour « Hebbian synapses with adaptive threshold », opérant dans les synapses glutamatergiques et n'utilisant qu'une règle hebbienne standard à deux termes.

Un certain nombre d'autres modèles des ganglions de la base considèrent les modèles TD comme acquis, les utilisent pour ajuster les poids des synapses corticostriatales, sans pour autant les mettre au centre de leur problématique. Cette utilisation est signalée dans les tableaux récapitulatifs 2.1 et 2.2 par la mention « Acteur/Critique ».

2.2.2 Mémoire à court terme et mémoire de travail

La mémoire à court terme désigne de façon générique un système de stockage temporaire de l'information, de capacité limitée. En général, la mémoire de travail est considérée comme la capacité à maintenir et à utiliser des représentations mentales pour les comportements orientés vers un but. Cela implique de posséder la capacité de sélectionner les informations à conserver et à extraire de la mémoire à court terme (Baddeley, 1993). Les modèles des ganglions de la base abordant les problèmes de mémoire à court terme et mémoire de travail ne font en général pas de distinction très nette entre les deux concepts et sont suffisamment proches dans leurs conceptions et fonctions pour qu'on les considère simultanément.

On considère en général que la mémoire à court terme se matérialise par l'activité persistante de neurones (plus particulièrement ceux du cortex préfrontal ou PFC) plutôt que par la modification de connexions synaptiques entre neurones, caractéristique de la mémoire à long terme. Alors que certain modèles considèrent que cette mémoire est localisée directement dans certains noyaux (striatum ou NST) des GB, la majorité des travaux considèrent qu'elle est stockée dans les zones antérieures du cortex. Ils conçoivent dans ce cas les GB comme le système de contrôle et de mise à jour de cette mémoire. Le rôle des GB est alors de sélectionner dans les entrées corticales celles qui sont significatives pour la tâche en cours de résolution, d'assurer leur maintien en mémoire et de supprimer les autres.

Localisation dans les ganglions de la base

Les travaux de Woodward *et al.* (1995) considèrent que si le striatum a une structure de réseau d'inhibitions latérales, il peut héberger une mémoire à court terme. En effet, de tels réseaux se stabilisent, sous l'effet d'excitations, dans des états dits ON/OFF, où un certain nombre de neurones sont fortement activés et les autres silencieux. Or ces états sont spécifiques des excitations les générant. Ils considèrent donc le striatum comme le lieu d'une mémoire à court terme dont les différentes valeurs sont les états ON/OFF. On notera que le modèle de génération de séquences de Berns et Sejnowski (1998), détaillé en 2.2.3, a la particularité d'utiliser une mémoire à court terme localisée dans le NST.

Localisation dans le cortex

Dans la seconde catégorie, qui localise cette mémoire dans le cortex frontal, on distinguera les modèles qui considèrent que l'activité des neurones est entretenue par : la boucle cortex-ganglions de la base-thalamus-cortex, la boucle cortex-thalamus-cortex ou des mécanismes d'autoexcitation des neurones corticaux.

1. *Boucle cortex-ganglions de la base-thalamus-cortex* :

Modèle	Domaine		Tâche	Chemins	Eléments modélisés					Rôle
	MCT	MdT			Ctx	GB	Th	CS	SNc	dopamine
Jackson & Houghton, 1992	●		Paradigme « attentional precuing »	Direct : mise à jour mémoire. Indirect : maintien mémoire	●	●	●		●	Module l'importance relative des deux chemins
Arbib & Dominey, 1992		●	Saccade occulaire retardée guidée par un indice	Direct uniquement	●	●	●	●	●	Apprentissage de la direction en fonction de l'indice fourni
Woodward et al., 1995	●		NA	NA		●				NA
Berns & Sejnowski, 1996	●		WCST, bandit multi-bras	Réseau off-center(direct) on-surround (indirect) ; Indirect : extinction de la sélection	●	●	●		●	Acteur/Critique
Gelfand et al., 1997	●		Sélection de réponse verbale	Direct uniquement	●	●	●			NA
Beiser & Houk, 1998	●		Encodage de séquences	Direct uniquement. Indirect envisagé pour la remise à zéro de la mémoire	●	●	●			NA
Berns & Sejnowski, 1998	●		Reproduction de séquences	Direct : sélection. Indirect : mémoire à court terme		●			●	Acteur/Critique
Monchi et al., 2000		●	WCST, cDRT, DMS	Direct uniquement	●	●	●			NA
Frank et al., 2000		●	Paradigme 1-2-AX	Direct uniquement	●	●	●			NA

TAB. *2.1: Tableau récapitulatif des modèles de mémoire à court terme / de travail. MCT : mémoire à court terme ; MdT : mémoire de travail ; Ctx : aires du cortex ; GB : ganglions de la base ; Th : noyaux thalamiques ; CS : colliculus supérieur ; SNc : substance noire compacte.*

Monchi *et al.* (2000) simulent l'exécution de divers tests standards : test de tri de cartes de Wisconsin (WCST), tâches de réponse retardée utilisant un indice sur la position de l'objet à pointer (cDRT) ou un indice donnant le type de l'objet (DMS) (voir l'annexe B pour les détails concernant ces tests). Les délais impliqués par l'ensemble de ces tests nécessitent l'utilisation d'une mémoire de travail. Elle est simulée par l'activité soutenue de neurones dans le PFC, entretenue par la propagation d'activité dans la boucle cortex-ganglions de la base-thalamus-cortex. Le modèle est à base de neurones artificiels, et, concernant ganglions de la base, seul le chemin direct (au sens d'Albin *et al.*) est modélisé. Il a pour rôle de sélectionner les neurones à activer via des inhibitions latérales situées dans le striatum et le GPi. Le déroulement du WCST ou l'alternance de tests cDRT et DMS dans une même session exigent des changements de stratégie de résolution de problème. L'originalité de ce modèle est donc de comporter plusieurs sous-boucles préfrontales, codant chacune une stratégie de résolution du problème et une boucle limbique qui, grâce aux informations concernant l'échec ou le succès de la tâche fournies par la connexion de l'amygdale sur le noyau accumbens, active l'une ou l'autre des stratégies implémentées dans les boucles préfrontales.

2. *Boucle cortex-thalamus-cortex* :

   L'existence de boucles cortico-thalamo-corticales est exploitée par certains modèles de mémoire à court terme pour expliquer l'activité soutenue de neurones du PFC. C'est le cas du modèle de Gelfand *et al.* (1997), utilisé pour simuler le phénomène d'inhibition proactive –une difficulté croissante, lors d'expériences successives, à se souvenir d'une série d'éléments appartenant à une même catégorie alors qu'un changement de catégorie rétablit les performances initiales. Là aussi, seul le chemin direct des GB est simulé et la sélection est réalisée par des connexions latérales inhibitrices dans le striatum. Le modèle d'Arbib et Dominey (Dominey et Arbib, 1992; Arbib et Dominey, 1995), qui simule la génération de saccades oculaires retardées, où l'information de direction donnée par un indice visuel doit être mémorisée temporairement, entre dans la même catégorie. Une fois de plus, seul le chemin direct est modélisé. Les ganglions de la base ont pour seul rôle de favoriser certaines positions de l'espace via un processus d'apprentissage sur les synapses cortico-striatales, la sélection proprement-dite de la direction de la saccade ayant alors lieu dans le colliculus supérieur. Enfin, bien que consacré à l'encodage de séquences et relevant donc également des modèles traités en 2.2.3, le modèle de Beiser et Houk (1998) fait aussi usage d'une mémoire à court terme entretenue par le rebouclage cortico-thalamo-cortical.

3. *Mécanismes d'autoexcitation des neurones corticaux* :

   Jackson et Houghton (Jackson et Houghton, 1992; Jackson et Houghton, 1995), bien que s'intéressant à la modélisation du contrôle de l'attention visuelle, proposent en réalité un modèle où le rôle des GB est de contrôler la mise à jour d'une mémoire à court terme. En effet, un système attentionnel –localisé dans le cortex frontal– y mémorise les zones du champ visuel où sont présents ou attendus des points de fixation. L'activité persistante nécessaire à cette mémorisation est induite par une excitation récurrente de chaque neurone sur lui-même. Les ganglions de la base créent, conservent ou suppriment ces auto-activations selon que les points de fixation effectivement observés correspondent ou non aux attentes. Ils sont modélisés par un réseau de neurones artificiels comportant les deux chemins (direct et indirect), structuré en canaux, où chaque canal représente une position dans le champ visuel. Le mécanisme de mise à jour de la mémoire est le suivant : les striosomes comparent l'activité du système attentionnel et du système perceptif et modifient le taux de dopamine dans les GB selon le degré de correspondance constaté (via leurs connexions vers la SNc). Les variations du niveau de dopamine favorisent soit le chemin indirect (inhibition de tous les canaux et donc conservation de l'état passé de la mémoire), soit le chemin direct (activation sélective des seuls canaux où attentes et perceptions correspondent, modification de l'état de la mémoire).

   Le modèle d'apprentissage par renforcement de Berns et Sejnowski (1996), présenté précédemment, utilise une mémoire de travail localisée dans le cortex préfrontal, où l'activité des neurones est maintenue par autoexcitation. Elle est constituée de plusieurs groupes de neurones, chacun intégrant des copies de la sortie des ganglions de la base sur des échelles de temps différentes (variant en fonction des poids des autoconnexions).

   Le modèle de Frank *et al.* (2000) considère que le maintien des informations en mémoire est assuré par une capacité intrinsèque des neurones corticaux à se maintenir actifs, qui reste à préciser sur un plan biologique. Le rôle des ganglions de la base est d'assurer une mise à jour rapide, une maintenance robuste (les informations de contexte doivent pouvoir par exemple être conservées indépendament du processus en cours) et une mise à jour sélective (elle doit pouvoir n'affecter que quelques éléments de cette mémoire) de la mémoire de travail. Le PFC est subdivisé en « bandes » spécialisées dans diverses catégories de données mémorielles (contexte comportemental, étape dans une séquence comportementale, actions possibles). Chacune de ces bandes voit certains de ses neurones activés, soit par une entrée sensorielle temporaire, soit par leur capacité d'autoexcitation. Les GB sélectionnent à chaque instant la catégorie d'informations (la bande) à mettre à jour, ce qui se traduit par le passage en activation maintenue intrinsèquement des neurones activés par les entrées sensorielles du moment et la remise à zéro des autres. Ce modèle a été testé

dans une tâche dite 1-2-AX (voir l'annexe B) où selon la succession d'indices « visuels » –lettres, chiffres– le système doit mémoriser le contexte et les séquences visuelles pour effectuer la bonne action.

Les modèles de mémoire de travail fondés sur une activité persistante de neurones dans le cortex préfrontal sont les plus biologiquement plausibles (Fuster, 1989; Goldman-Rakic, 1994). La plupart d'entre eux attribuent aux GB le rôle de sélectionner les informations à stocker en mémoire à court terme ou à supprimer de cette mémoire. Enfin, tous ces modèles mémorisent explicitement chaque information dans un neurone cortical spécifique, à l'exception de celui de Berns et Sejnowski (1996), qui propose une notion plus floue de la mémoire, où l'information stockée est intégrée à diverses constantes de temps.

2.2.3 Génération de séquences

La mémoire à court terme et la génération de séquences sont souvent liées, la première servant de base à la seconde. On a ainsi vu que le modèle de mémoire de travail de Frank *et al.* (2000) permet, entre autres, la génération de séquences. On constatera que les modèles liés à la génération de séquence présentés ici utilisent des mémoires à court terme. La génération de séquences, qu'il s'agisse de mots, de mouvements ou de pensées, nécessite deux capacités principales, tout d'abord, celle d'apprendre et donc de stocker ces séquences, puis celle de les restituer sans altération.

Apprentissage de séquences

Le modèle de Beiser et Houk (1998), une implémentation du modèle conceptuel proposé par Houk et Wise (1995), correspond à la première étape : l'encodage d'une séquence temporelle dans la structure spatiale du réseau de neurones du cortex préfrontal. Comme évoqué précédemment, l'activation persistante est obtenue par une boucle excitatrice entre les neurones du PFC et ceux du thalamus. Seul le chemin direct des ganglions de la base est modélisé par des neurones artificiels organisés en canaux parallèles. Les projections corticales ont lieu sur l'ensemble des neurones du striatum suivant des poids aléatoires, le striatum est le lieu d'inhibitions –soit latérales, soit par des interneurones– et se projette ensuite via le GPi sur le thalamus en autant de canaux qu'il y a de paires de neurones corticaux et thalamiques affectées à la mémoire. L'activité sensorielle générée par l'apparition d'un nouvel élément de la séquence en cours d'encodage vient se superposer à l'activité déjà engendrée par les neurones mémoriels préfrontaux activés. L'activité résultante génère alors la désinhibition sélective d'un nouveau neurone préfrontal. L'itération de ce processus aboutit à l'activation persistante d'un groupe de neurones du PFC spécifique à la séquence. Le point fort de ce modèle est que, si ses poids

Modèle	Domaine		Tâche	Chemins	Eléments modélisés				Rôle
	ES	RS			Ctx	GB	Th	SNc	dopamine
Beiser & Houk, 1998	●		Encodage de séquences	Direct uniquement. Indirect envisagé pour la remise à zéro de la mémoire	●	●	●		NA
Berns & Sejnowski, 1998	●	●	Reproduction de séquences	Direct : sélection. Indirect : mémoire à court terme		●		●	Acteur/Critique
Bischoff, 1998		●	Génération de séquences de mouvement d'un bras	Direct : prévision du prochain état. Indirect : extinction du mouvement	●	●	●	●	Module l'importance relative des deux chemins
Hikosaka et al., 1999	●	●	Tâche 2x5	NA	●	●	●	●	Acteur/Critique à deux vitesses

TAB. *2.2: Tableau récapitulatif des modèles d'apprentissage et de restitution de séquences. ES : encodage de séquence ; RS réstitution de séquence ; Ctx : aires du cortex ; GB : ganglions de la base ; Th : noyaux thalamiques ; SNc : substance noire compacte.*

aléatoires sont initialisés dans certaines gammes de paramètres, l'obtention de ces motifs d'activation différents selon les séquences d'entrée ne nécessite aucun processus d'adaptation, c'est une propriété inhérente à sa structure.

Restitution de séquences

Le modèle de Bischoff (1998) est consacré à l'exécution de séquences de mouvements des membres déjà mémorisées. Il suppose que la mémoire de travail du PFC est en position de déclencher une mémoire de séquence dans l'aire prémotrice supplémentaire, qui charge ensuite les GB de générer la séquence voulue. Bischoff considère que les ganglions de la base jouent un rôle double, incarné par les deux chemins, direct et indirect. Le chemin direct estime le prochain état sensoriel, en fonction de l'action en cours, ce qui permet de générer des séquences par comparaison avec l'activité de l'aire prémotrice supplémentaire. Le chemin indirect, lui, est chargé de l'extinction des mouvements, par exemple en fin de séquence, par un renforcement de l'excitation diffuse du NST sur le globus pallidus interne (équivalent chez le primate du noyau entopédonculaire) et la substance noire réticulée, ce qui inhibe le cortex moteur.

Apprentissage et restitution de séquences

Le modèle d'apprentissage et de restitution de séquences de Berns et Sejnowski (1998), évoqué en 2.2.2, est unique en ce qu'il considère une mémoire à court terme sise dans les poids des connexions de la boucle NST-GPe. En effet, d'un côté, le chemin direct des GB est supposé effectuer une sélection entre divers canaux, pouvant représenter des actions. De l'autre, le chemin indirect sert à apprendre et à reproduire les séquences présentées au système. Cet apprentissage prend la forme d'un apprentissage TD classique, si ce n'est que le renforcement ne correspond pas ici à des récompenses issues de l'environnement, mais au calcul interne de l'écart entre les séquences présentées et générées, sans considération pour l'efficacité de ces séquences. Ce système est capable d'apprendre des séquences et de les générer si on lui en présente le début ou une version dégradée. On notera l'intérêt porté ici aux connexions dopaminergiques touchant le globus pallidus externe et interne (équivalents chez les primates du globus pallidus et du noyau entopédonculaire) peu documentées et rarement reprises dans les modèles, ainsi que l'utilisation d'hypothétiques projections du globus pallidus interne sur la SNc.

Nakahara, Hikosaka *et al.* (Hikosaka *et al.*, 1999; Nakahara *et al.*, 2001) étudient les bénéfices que l'on peut tirer de l'utilisation de deux boucles cortex-ganglions de la base-thalamus-cortex concurrentes pour l'apprentissage de séquences de mouvement des bras (tâche 2x5, voir annexe B). En effet, leur système intègre une première boucle préfrontale, disposant d'une mémoire de travail et d'un fort taux d'apprentissage lui permettant d'acquérir rapidement la séquence, mais souffrant d'un manque de précision dans la commande motrice. La seconde boucle, motrice, qui n'est pas dotée de mémoire de travail, apprend plus lentement, mais commande le bras avec précision et permet des mouvements plus rapides. L'apprentissage utilise un modèle par renforcement standard. L'aire prémotrice supplémentaire coordonne l'action des deux boucles, favorisant la préfrontale lors de la découverte d'une nouvelle séquence, puis la motrice lorsque la résolution de la tâche passe au domaine des habitudes.

2.2.4 Contrôle de trajectoire bas niveau

Les modèles de mémoire à court terme et de génération de séquences font en général référence à la boucle préfrontale. Nous allons ici aborder deux modèles consacrés à la boucle motrice, puisque s'intéressant au contrôle de la trajectoire des membres via la commande directe des mouvements articulaires. Ils sont assez atypiques, puisque dérivés de travaux de robotique classique utilisant la métaphore des « grilles résistives ». Ces grilles représentent une discrétisation de l'espace des configurations possibles sous forme d'un ensemble de nœuds reliés entre eux par des résistances. Les obstacles y sont modélisés par des noeuds de conductivité nulle et les points « origine du mouvement » et « cible à atteindre » par des potentiels appliqués aux

Modèle	Tâche	Chemins	Eléments modélisés				Rôle dopamine
			Ctx	GB	Th	SNc	
Connolly & Burns, 1993	Simulation de mouvements de bras et jambes robotiques	NA		●			Activation dans le striatum des neurones représentant l'état courant
Lörincz, 1997	Simulation de mouvements de bras robotique	Intégration de la trajectoire commandée (direct) et effective (indirect)	●	●			NA

TAB. *2.3: Tableau récapitulatif des modèles de contrôle de trajectoire. Ctx : aires du cortex ; GB : ganglions de la base ; Th : noyaux thalamiques ; SNc : substance noire compacte.*

nœuds de la grille correspondants. Le calcul du champ de potentiel sur l'ensemble des points de la grille permet de choisir simplement le prochain déplacement par une descente de gradient.

Dans le cas du modèle de Connolly et Burns (1993, 1995), seul le striatum est modélisé. Supposant l'existence d'inhibitions latérales entre les neurones du striatum, ils l'assimilent à la grille résistive. La plausibilité biologique de ce modèle est mise en doute, en particulier à cause du rôle prépondérant des inhibitions latérales du striatum (Houk *et al.*, 1995b; Bischoff, 1998).

Le modèle de Lörincz (Lörincz, 1997) considère que la grille résistive est située dans le cortex, le rôle des ganglions de la base étant limité à la soustraction de la commande issue d'une première grille et de la position réelle issue d'une seconde grille, par une rapide assimilation de cette opération au rôle des chemins direct et indirect. L'absence de modélisation des noyaux des GB et l'exploitation d'une seule de leurs caractéristiques limitent là aussi grandement l'aspect biomimétique du modèle.

Il apparaît que de nouveaux modèles des boucles motrices des GB, plus proches des données biologiques et délimitant plus clairement ce qui est de leur domaine et ce qui est de celui du cervelet, doivent être développés (Doya, 2000).

2.3 Bilan

D'un point de vue modélisation, les modèles passés en revue ont deux limitations principales :

– En dehors de ceux ne s'attachant qu'à la modélisation du striatum, ils sont fondés sur l'interprétation de la connectivité des GB en terme de chemin direct/chemin indirect proposée par (Albin *et al.*, 1989) et nombre d'entre eux se contentent de modéliser le seul chemin direct. Cette interprétation a cependant vieilli, de l'aveu même de ses auteurs (Albin *et al.*, 1995) et n'est plus en mesure de traduire fidèlement les connaissances actuelles

concernant les ganglions de la base.

– De la même façon, la majeure partie d'entre eux voient leur fonctionnement centré sur l'effet sélectif d'inhibitions latérales dans le striatum, inhibitions dont l'existence est soumise à controverse (Jaeger *et al.*, 1994).

Un modèle récemment mis au point Gurney *et al.* (2001a) lève ces limitations. En effet, il propose une réinterprétation de la proposition d'Albin *et al.* (chemins direct et indirect), mettant en valeur deux sous-circuits symétriques des GB, l'un de *sélection* et l'autre de *contrôle de la sélection*. Il s'avère plus proche des données anatomiques et électrophysiologiques que ses prédécesseurs, et a déjà donné lieu à des expériences de sélection de l'action *situées* à bord d'un robot (Montes-Gonzalez *et al.*, 2000). Développant les idées présentées dans (Redgrave *et al.*, 1999b), il considère que la dopamine joue un rôle de facilitation de la transition entre comportements, et non de signal d'apprentissage. C'est pourquoi ce modèle est actuellement dénué de capacités d'apprentissage.

C'est à partir de ce modèle que nous implémenterons l'architecture de sélection de l'action d'un robot dépourvu de système de navigation. Elle sera testée pour la première fois dans une tâche de survie. L'adaptation du modèle ainsi que les expériences correspondantes sont décrites dans le chapitre suivant.

Les modèles d'apprentissage par renforcement, eux, ont des limitations qui ne leur permettent pas, dans l'état actuel, de résoudre le problème de la sélection de l'action. Une exception serait le modèle proposé par Dayan (2001). Il cherche à concilier l'apprentissage par renforcement –qui est adapté à la résolution d'une tâche motivée par un but unique– et la sélection de l'action –qui s'intéresse également à l'influence de motivations contradictoires sur le choix du but courant– en adjoignant aux modèles standards un module motivationnel. Il considère pour cela les circuits ventraux des ganglions de la base, dont le noyau d'entrée (le noyau accumbens) est subdivisé en deux régions (le « core » et le « shell »). Il associe le « shell » au choix d'une motivation et le « core » au choix de l'action. Bien qu'encore incomplet, selon l'auteur, ce modèle ouvre une voie intéressante, considérant le rôle du noyau accumbens et de la boucle limbique (ventrale) des ganglions de la base et permettant de concilier apprentissage par renforcement et sélection de l'action.

Nous évoquerons à nouveau dans le chapitre 4 le rôle éventuel des circuits ventraux dans la sélection d'une stratégie de résolution de problèmes proposé par Monchi *et al.* (2000).

Chapitre 3

Modèle biomimétique de sélection de l'action

To make a model which would reproduce all the behavior of a rat would require a mechanism probably as large as the Capitol in Washington
Hull (1935)

3.1 Le modèle de Gurney, Prescott et Redgrave

3.1.1 Une nouvelle interprétation des GB

Le modèle développé par Gurney, Prescott et Redgrave (Gurney *et al.*, 2001a, 2001b) –que l'on nommera GPR, d'après les initiales de ses auteurs– à l'« Adaptive Behaviour Research Group » (ABRG) de l'Université de Sheffield, propose une révision du schéma chemin direct/chemin indirect d'Albin *et al.*, qui domine la conception des modèles computationnels des ganglions de la base. Cette interprétation classique de la connectivité des ganglions de la base considère un chemin inhibiteur direct depuis le striatum vers l'EP/SNr et un chemin indirect excitateur dans lequel le striatum inhibe le globus pallidus, levant l'inhibition tonique que ce dernier exerce sur le noyau subthalamique qui peut alors exciter l'EP/SNr (fig 3.1, gauche). En effet, ce dernier a l'inconvénient de négliger un certain nombre de connexions importantes des GB (voir fig. 3.1, gauche) :

- les nombreux afférents corticaux sur le noyau subthalamique,
- les efférences inhibitrices du noyau subthalamique sur le globus pallidus,
- les projections directes du globus pallidus sur l'EP/SNr.

Dès 1995, Parent et Hazrati (1995a, 1995b) mettent en doute le schéma d'Albin *et al.*, en attribuant, d'une part, un rôle de contrôle du GP sur l'EP, la SNr et le TRN et, d'autre part, un rôle

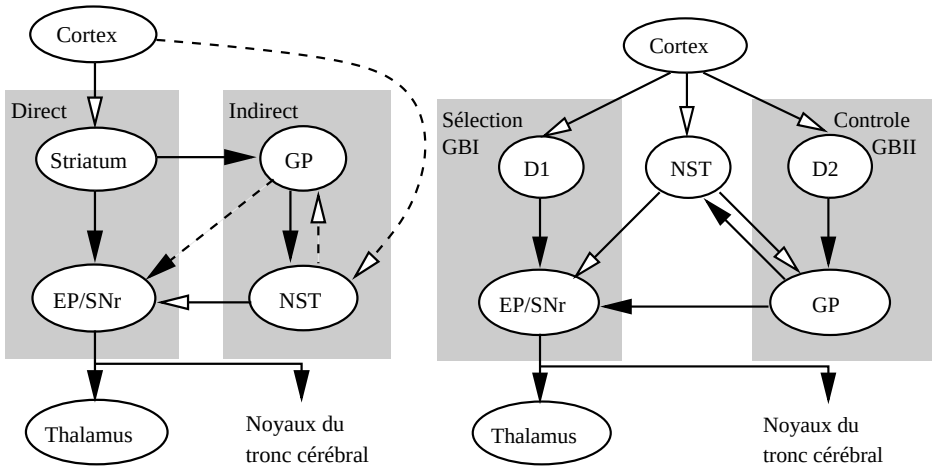

FIG. *3.1: Interprétations de la connectivité des ganglions de la base. Gauche : modèle d'Albin et al., chemin direct / chemin indirect. Droite : modèle GPR, module de sélection / module de contrôle. Flèches évidées : connexions excitatrices. Flèches pleines : connexions inhibitrices. Flèches en pointillés : connexions négligées.*

NST sur le niveau d'excitation général des GB. Le GPR constitue une nouvelle interprétation de la connectivité des GB, dont la clef est la distinction de deux populations de neurones du striatum –possédant des récepteurs dopaminergiques de type D1 ou D2– se projetant respectivement sur l'EP/SNr et le GP. Cette distinction met à jour deux modules symétriques –*GBI* et *GBII*– se partageant le NST (voir fig. 3.1, droite). Cette interprétation considère que le rôle générique des ganglions de la base est d'effectuer une sélection de type « qui perd gagne » sur tous types de signaux.

3.1.2 Fonctionnement du modèle

Entrées

Le modèle GPR a notamment été appliqué au problème de la sélection de l'action tel qu'il a été défini au chapitre 1. Dans ce cadre, les signaux sur lesquels il opère sont des variables spéculatives nommées « saliences » qui mesurent la nécessité d'effectuer un comportement donné à un instant donné. Ces saliences –qui pourraient s'assimiler aux FFC des éthologues (cf. 1.2.1)– permettent de mettre en compétition les comportements en les évaluant sur un critère comparable, par une « common currency » (McFarland et Sibly, 1975).

Le calcul des saliences est supposé être distribué entre les zones du cortex se projetant sur le striatum et le striatum lui-même. Il consiste en une somme pondérée des informations externes et internes en rapport avec le comportement évalué. Les vecteur de poids (W_i^P, W_i^E, W_i^I) correspondants (voir fig. 3.2) sont ajustés à la main. Le résultat de ce calcul est fourni

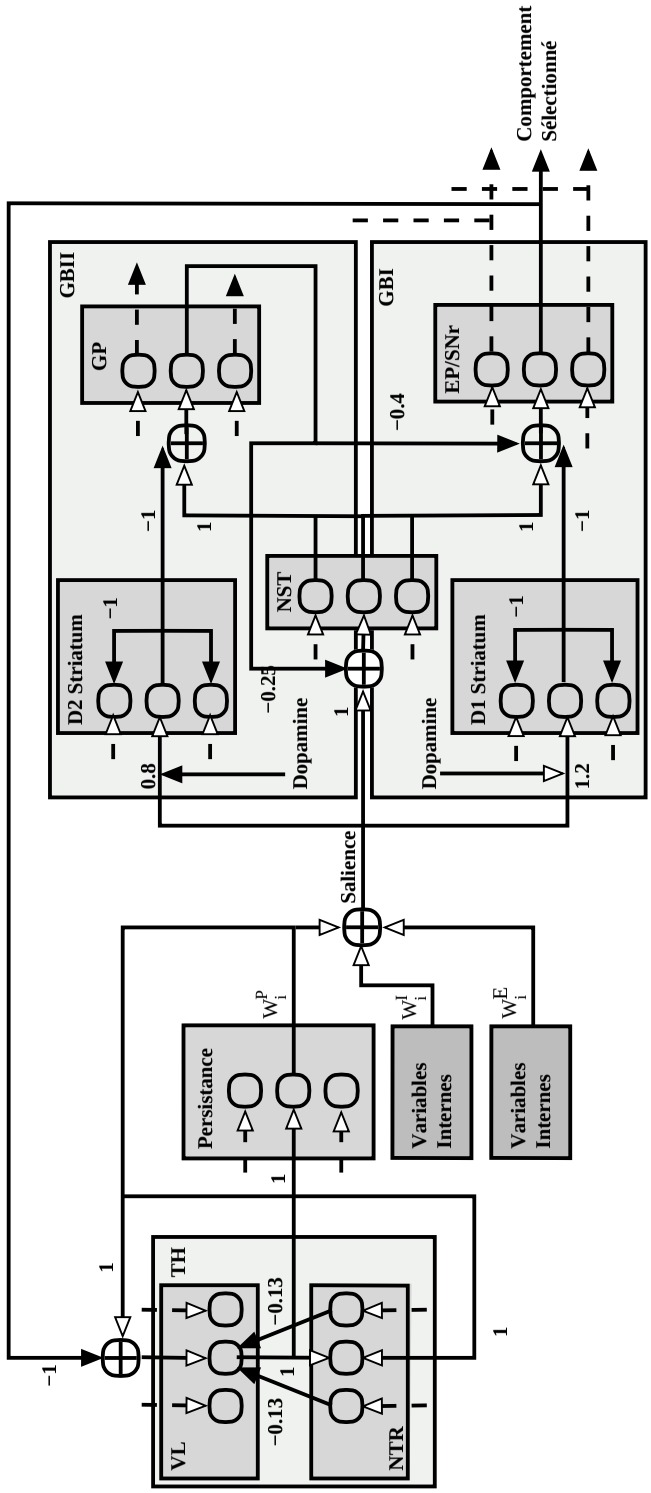

FIG. *3.2: Schéma du modèle GPR, exemple à 3 canaux. Détails dans le texte. Flèches évidées : connexions excitatrices. Flèches pleines : connexions inhibitrices.*

aux modules d'entrée du système (striatum D1 et D2, NST). On notera là une simplification du modèle, qui fournit en entrée du NST les mêmes valeurs qu'en entrée du striatum, alors que ce même striatum est supposé être partie prenante du calcul des saliences, et que le NST reçoit des projections corticales émanant de neurones n'appartenant pas à la même couche que ceux se projetant sur le striatum.

Connexions internes

A l'instar du modèle de Berns et Sejnowski (1996), les signaux (ici saliences) en compétition passent par les canaux des GB. Chaque canal est matérialisé, dans chaque noyau, par un neurone artificiel de type intégrateur à fuite (voir annexe C.1), qui est le plus simple modèle neural incorporant la notion de potentiel de membrane dynamique (Arbib, 1995; Yamada *et al.*, 1989).

Sélection

Le module *GBI*, qui commande directement les sorties des GB, réalise la tâche de *sélection* proprement dite par le biais de deux mécanismes.

Le premier concerne des circuits inhibiteurs récurrents au sein du striatum D1, qui tendent à sélectionner un signal d'entrée gagnant –le plus élevé– et à générer en sortie de ce noyau un unique signal non-nul dans le canal gagnant. Ce signal de sortie est transmis à l'EP/SNr sous forme d'une entrée inhibitrice. Les neurones de l'EP/SNr sont toniquement actifs et envoient donc un flot d'inhibitions continuel sur les cibles des ganglions de la base. Lorsque le signal issu du striatum D1 inhibe un canal spécifique de l'EP/SNr, il lève alors sélectivement l'inhibition exercée sur une sous partie des cibles des ganglions de la base.

Le second mécanisme de sélection est un réseau « feed-forward » de type « off-center on-surround », semblable à celui utilisé par Berns et Sejnowski (1996). L'effet « on-surround » y est fourni par l'excitation diffuse du noyau subthalamique sur tous les canaux de l'EP/SNr, et l'effet « off-center » par le signal inhibiteur du striatum D1. Ce mécanisme sert à renforcer les différences –a améliorer le contraste– entre le canal de l'EP/SNr gagnant et les autres.

Contrôle

Ce circuit de sélection est modulé par des signaux de *contrôle* issus du circuit *GBII*, dont l'architecture est similaire à celle du *GBI* : inhibitions récurrentes locales dans le striatum D2 et un réseau de type « off-center on-surround », dans lequel l'effet « on-surround » est fourni par l'excitation diffuse de tous les canaux du GP par le NST, et l'effet « off-center » par le signal inhibiteur du striatum D2. Les sorties du GP inhibent les canaux de l'EP/SNr et du NST. Le rôle de ce circuit est triple.

Premièrement, la sortie du GP dirigée vers l'EP/SNr renforce la sélectivité entre les canaux.

Deuxièmement, l'inhibition exercée par le GP sur le NST sert à réguler automatiquement l'activité du NST, ce qui permet une sélection efficace indépendante du nombre de canaux impliqués. Sans ce mécanisme, une augmentation du nombre de canaux génère une augmentation de l'excitation issue du NST, jusqu'au point où les inhibitions issues du striatum D1 ne peuvent plus la contrebalancer et où tous les canaux sont donc rendus inactifs. La rétroaction négative du GP est juste suffisante pour ajuster automatiquement l'excitation du NST de sorte que la sélection fonctionne toujours.

Le troisième rôle du *GBII* concerne la modulation exercée par les récepteurs D2 qui procède en synergie avec les récepteurs D1 pour renforcer encore le contraste entre canaux. En effet, l'action de la dopamine est modélisée très simplement par une augmentation de la transmission corticale vers les neurones D1 (sélection) et une diminution équivalente vers les neurones D2 (contrôle). Le réglage de ce déséquilibre permet de modifier les aptitudes du modèle à sélectionner nettement un canal ainsi qu'à changer de canal sélectionné.

Boucle Thalamo-corticale

Le modèle constitué des deux modules *GBI* et *GBII*, se voit adjoindre un circuit thalamo-cortical (*TH*), complétant la boucle cortex-GB-thalamus-cortex. Il est composé du noyau thalamique réticulé (NTR), du thalamus ventro-latéral (VL) et de neurones du cortex (Humphries et Gurney, 2002). Tous ces modules ont les mêmes canaux ségrégés et modélisés par des neurones intégrateurs à fuite que le *GBI* et le *GBII*. Ce circuit constitue une boucle de rétroaction positive fournissant au signal d'entrée de chaque canal un « bonus », nommé *persistance*, d'autant plus fort que le canal en question est sélectionné.

Sorties

La sélection effectuée par les ganglions de la base fonctionne sur le principe de la désinhibition. Dans le cas où le GPR est soumis à une entrée constante, ses sorties se stabilisent, de sorte que le module EP/SNr inhibe uniformément les canaux de sortie à l'exception de celui qui est sélectionné. Cependant, dans la pratique, les entrées du GPR varient de façon continue, et il arrive fréquemment que plusieurs comportements soient partiellement désinhibés (c'est-à-dire soumis à une plus faible inhibition que la majorité des autres). Dans ce cas, les sorties du sytème sont des ordres moteurs envoyés aux effecteurs qui sont la résultante de la somme des ordres générés par chaque comportement, pondérée par leur degré de désinhibition au sortir du GPR (méthode de sélection dite de « soft-switching »).

GPR et mécanisme « winner-takes-all »

Le mécanisme principal de sélection d'un GPR est un réseau de neurones artificiels de type « off-center on-surround », c'est à dire une variante de mécanisme de type « gagnant prend tout » (« winner-takes-all » ou WTA), adaptée à la contrainte de sélectionner par désinhibition. De ce fait, le rôle joué par le modèle GPR peut paraître similaire, dans un premier abord, à celui d'un WTA excessivement complexe. Cependant, la présence d'une boucle de rétroaction positive et d'un mécanisme de contrôle de l'activité interne comprenant également une boucle lui adjoint des propriétés dynamiques supplémentaires.

3.1.3 Expérimentations réalisées à l'ABRG

Le GPR a d'abord été testé en simulation, isolé de tout environnement et de toute tâche à accomplir, ceci afin de s'assurer qu'il exhibe des propriétés de transitions rapides, d'absence de distortion et de persistance, nécessaires pour un mécanisme de sélection de l'action (Gurney *et al.*, 2001b).

La première expérience robotique d'évaluation du modèle GPR en terme de sélection de l'action a été menée par Montes-Gonzalez (Montes-Gonzalez *et al.*, 2000; Montes-Gonzalez, 2001), sur un robot Khepera (©K-Team). Il s'agissait, d'une part, de reproduire les séquences comportementales d'un rat en manque de nourriture, placé dans une arène carrée vivement éclairée au centre de laquelle se trouve de la nourriture et dont seul l'un des coins est à l'ombre. D'autre part, les effets de la modulation du niveau de dopamine sur les transitions comportementales étaient testées durant cette tâche.

Sous l'effet supposé de deux motivations correspondant à la peur et la faim, le rat effectue dans l'arène un enchaînement d'actions stéréotypé : il commence par longer les murs jusqu'au coin sombre, où il reste un certain temps, puis il va chercher de la nourriture au centre, qu'il rapporte dans le coin sombre et qu'il consomme.

Le robot Khepera (©K-Team), équipé d'une pince, est placé de la même façon dans une arène carrée au centre de laquelle se trouvent quatre cylindres de bois représentant la nourriture. Les quatre coins de l'arène sont équivalents. Il possède cinq comportements : *AllerVersMur*, *LongerMur*, *RechercherCylindre*, *PrendreCylindre*, *Lâcher Cylindre*. Les saliences de ces comportements sont calculées à partir de 4 variables externes et 2 variables internes. Les variables externes sont des booléens dérivés de mesures des senseurs infrarouges de proximité et du senseur de contact de la pince du robot. Il s'agit de *MurDétecté*, *CoinDétecté*, *CylindreDétecté* et *ObjetTenu*. Les deux variables internes sont la peur et la faim. La peur, initialement forte, diminue continûment au cours du temps. De son côté, la faim, initialement forte, ne diminue que lorsque le robot attrape un cylindre et le lâche hors de l'arène.

Le comportement du robot mime celui du rat : tant que la peur est élevée, il tend à se

diriger vers les coins en activant successivement *AllerVersMur* et *LongerMur*. Lorsque la peur a assez baissé, la faim prend le dessus et déclenche *RechercherCylindre* et *PrendreCylindre* pour trouver des cylindres, puis *AllerVersMur*, *LongerMur* et *Lâcher Cylindre* pour déposer les cylindres hors de l'arène au niveau des coins (ce qui correspond à la consommation de nourriture).

L'étude des effets engendrés par la manipulation du niveau de dopamine dans le modèle montre qu'en cas de déficit de dopamine, le robot a du mal à totalement désinhiber les comportements, ce qui entraîne un ralentissement dans leur exécution, que Montes-Gonzalez rapproche de la bradykinesie des patients parkinsoniens. D'autre part, si le niveau de dopamine est au contraire plus élevé que la normale, il arrive que le robot sélectionne deux comportements en simultané, ce qui génère des mouvements inappropriés (par exemple, le robot alterne de façon répétée *PrendreCylindre* et *Lâcher Cylindre* alors qu'il est en train de parcourir l'arène en activant *RechercherCylindre*), qu'il rapproche là du syndrome de Tourette.

3.2 Evaluation dans une tâche de survie

L'expérimentation robotique de Montes-Gonzalez analyse l'efficacité du GPR en tant que système permettant des transitions efficaces entre comportements dans une tâche plus simple que celle qui est classiquement réalisée en sélection de l'action car elle ne comporte pas de contrainte de survie. C'est pourquoi nous avons proposé une évaluation de ce modèle dans un problème de survie à deux ressources, qui constitue le scénario minimal pour ce type de tâche (Spier et McFarland, 1996).

Il s'agit pour le robot de sélectionner efficacement ses comportements afin d'assurer sa survie, en maintenant ses variables d'état interne dans des intervalles tolérables, sa *zone de viabilité* (Ashby, 1952). Cette survie dépend directement de la capacité de l'animat à se ravitailler auprès de deux types de ressources différents, en un temps limité par son niveau de recharge. L'utilisation de deux ressources différentes force l'animat à se déplacer dans l'environnement pour accéder à l'une puis à l'autre et le met en situation de conflit pour déterminer quelle ressource est prioritaire à un instant donné, susceptible de générer des oscillations comportementales.

L'évaluation du modèle consistera à comparer ses performances avec un modèle WTA minimal, dénué de boucles de rétroaction (Girard *et al.*, 2002; Laithier *et al.*, 2002; Girard *et al.*, 2003a).

Une première expérience comparera les performances de survie du GPR et du WTA afin de répondre aux questions suivantes :

- l'un des deux systèmes a-t-il des capacités à survivre plus longtemps que l'autre ?
- les variables internes sont-elles maintenues dans les plages de valeurs les plus confor-

tables –les plus éloignées des bornes des intervalles– de la *zone de viabilité* ?
– en quoi les différences d'architecture entre le GPR et le WTA affectent-elles leurs stratégies comportementales (durées et fréquences d'activations des comportements) ?

Nous avons ensuite cherché à analyser la qualité de la sélection des actions exercée par le GPR, en particulier ses capacités liées à la production de persistance comportementale. Une seconde expérience mettra donc les deux systèmes, GPR et WTA, en situation de conflit entre deux comportements à salience forte, afin de comparer leurs capacités à limiter leur oscillations comportementales.

La troisième expérience testera les capacités des deux systèmes à exploiter un comportement à faible coût énergétique pour limiter leur dépense énergétique et maintenir leurs variables internes dans des *zones de confort*.

3.2.1 Matériel et Méthodes

La tâche générale que le robot aura à effectuer sera de survivre dans un environnement où il pourra trouver deux types de ressources : des zones d'« ingestion » qui lui permettront de faire des réserves et des zones de « digestion » où il pourra « assimiler » ses réserves et les transformer en énergie utilisable. Sachant que tous les comportements du robot consomment de l'énergie, il va donc devoir alterner phases d'« ingestion » et de « digestion » pour survivre.

Environnement

L'environnement expérimental est une surface plane de 2m x 1,60m ceinte de murs de 10cm de hauteur (fig. 3.3, gauche). Elle est recouverte de carreaux de 40cm $\times$ 40cm, de trois types différents : 16 carreaux uniformément gris (un gris neutre représentant les zones neutres), 2 carreaux couverts d'un gradient circulaire allant du gris vers le noir (zones d'« ingestion »), dont les ressources sont inépuisables, et 2 carreaux couverts d'un gradient circulaire allant du gris vers le blanc (zones de « digestion »).

Variables externes

Le robot est entièrement réalisé avec des pièces standard des Lego Mindstorms (fig. 3.3, droite). On mesure 4 variables externes : *Blancheur*, (L_B), *Noirceur* (L_N), *ContactGauche* (C_G) et *ContactDroit* (C_D). Elles sont fournies par quatre senseurs.

Le robot possède deux capteurs de luminosité, disposés l'un derrière l'autre et dirigés vers

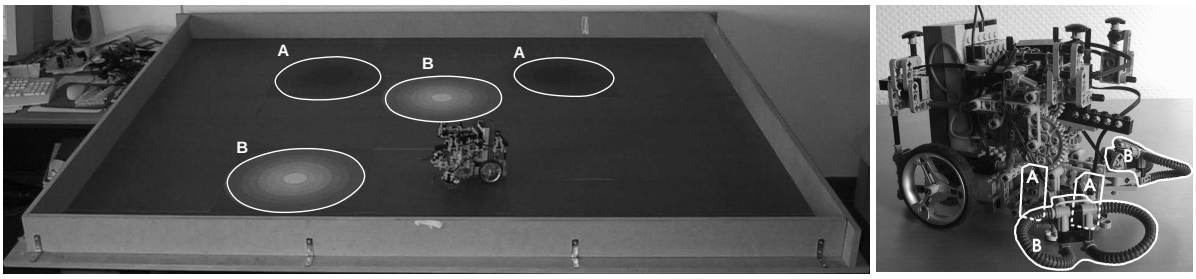

FIG. *3.3: Le robot Lego et son environnement. Gauche : Environnement ; (A) zones d'ingestion ; (B) zones de digestion. Droite : Robot ; (A) capteurs de luminosité ; (B) capteurs de contact.*

le sol (fig. 3.3, droite). La moyenne filtrée[1] des deux valeurs produites par ces capteurs sert à calculer deux variables, *Blancheur* (L_B) et *Noirceur* (L_N). La *Blancheur* (respectivement *Noirceur*) vaut 0 pour toute valeur de teinte plus sombre (respectivement claire) que le gris neutre et augmente linéairement pour les teintes plus claires (respectivement sombres), atteignant 1 pour les zones centrales blanches (respectivement noires).

Deux capteurs de contact, situés à l'avant-gauche et l'avant-droite, produisent chacun une valeur binaire (C_G et C_D) qui vaut 1 lorsque le robot entre en contact avec un obstacle.

Variables internes

Le robot possède trois variables internes : *Saleté*, *Energie Potentielle* et *Energie*. Elles sont toutes les trois à valeurs entre 0 et 255 (elles sont normalisées à des valeurs entre 0 et 1 avant leur utilisation pour le calcul des saliences).

– l'*Energie Potentielle* (E_{Pot}) correspond aux réserves puisées sur les zones d'« ingestion ». Lors de l'activation du comportement de recharge correspondant (*RechargeSurNoir*, voir plus bas), sur une zone sombre, le gain en E_{Pot} (ΔE_{Pot}) est proportionnel à la durée d'activation du comportement, T_{ingest} (en s), et à la *Noirceur* du sol :

$$\Delta E_{Pot} = 7 \times T_{Ingest} \times L_N \tag{3.1}$$

– l'*Energie* (E) est l'énergie réellement utilisable pour survivre dans l'environnement, elle est obtenue par « digestion » de l'E_{Pot} sur les zones claires. Pour survivre, le robot doit maintenir E au dessus de 0. Lors de l'activation du comportement de recharge correspondant (*RechargeSurBlanc*, voir plus bas), sur une zone claire, l'*Energie Potentielle* consommée (ΔE_{Pot}) et l'*Energie* générée (ΔE) sont proportionnelles à la durée d'activation du comportement T_{Digest} (en s) et à la *Blancheur* du sol :

[1] filtre médian appliqué aux 7 dernières valeurs.

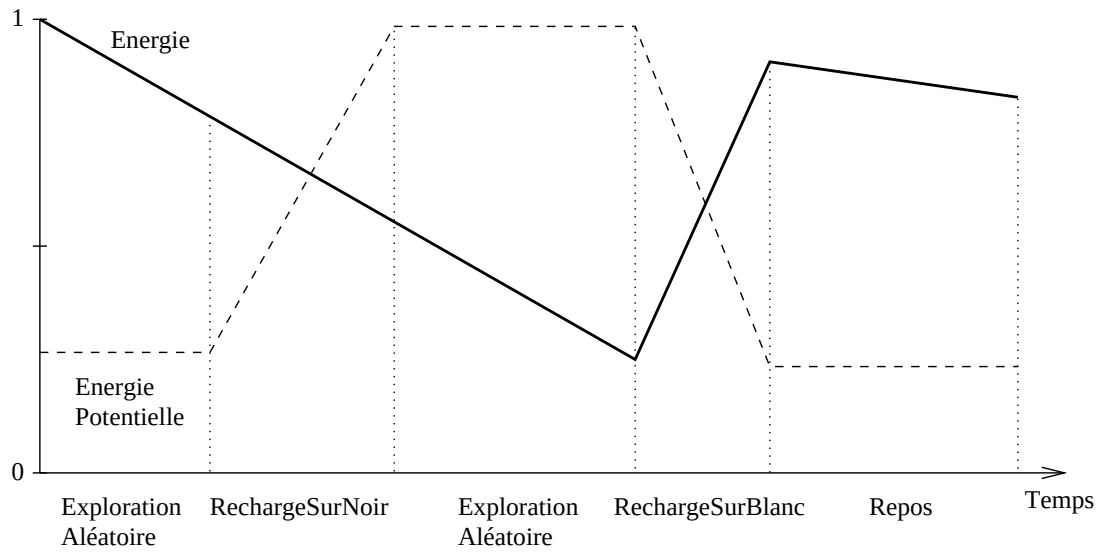

FIG. *3.4: Variations de l'*Energie Potentielle *et de l'*Energie *durant une séquence-type de comportements. On constate qu'en dehors des phases de* RechargeSurBlanc, *l'*Energie Potentielle *est constante, et que le taux de consommation d'*Energie *lors des phases de* Repos *est deux fois moindre que lors des autres comportements.*

$$\Delta E_{Pot} = -7 \times T_{Digest} \times L_B \tag{3.2}$$

$$\Delta E = T_{Digest} \times (7 \times L_B - 0,5) \tag{3.3}$$

Il en découle que lorsque la variable E_{Pot} n'est pas nulle, l'activation de *RechargeSurBlanc* produit effectivement de l'E, alors que lorsqu'il n'y a plus d'E_{Pot} disponible, l'E est consommée à raison de 0,5 unité.s^{-1}.

– La *Saleté* (S), n'est utilisée que pour l'expérience 2 et n'a pas d'effet direct sur la survie du robot. Elle augmente en permanence à un taux de 1 unité.s^{-1}, et peut être diminuée au taux de 4 unité.s^{-1} par l'activation du comportement *Toilette*.

Le maintien du robot dans sa *zone de viabilité* à proprement parler consiste uniquement à conserver son *Energie* au dessus de 0, même si la conservation de réserves d'*Energie Potentielle* est nécessaire pour atteindre cet objectif. La notion de *zone de confort* concerne la capacité à maintenir les variables internes éloignées des frontières de la *zone de viabilité*, ce qui dans notre cas consiste à maintenir l'*Energie* et incidemment l'*Energie Potentielle* aux plus hautes valeurs possibles.

Comportements

Selon les expériences, le robot est doté de comportements parmi les suivants :

– *RechargeSurNoir (RSN)* : le robot s'arrête, s'il est sur un emplacement plus sombre que celle du gris neutre, il « ingère » de la nourriture virtuelle, c'est-à-dire qu'il se recharge en *Energie Potentielle* qu'il ne peut utiliser immédiatement.
– *RechargeSurBlanc (RSB)* : le robot s'arrête, s'il est sur un emplacement de teinte plus claire que le gris neutre, il « digère » la nourriture précédemment « ingérée », c'est-à-dire qu'il transforme l'*Energie Potentielle* disponible en *Energie* utilisable pour exécuter ses comportements.
– *ExplorationAléatoire (EA)* : le robot se déplace aléatoirement (alternance de mouvements de translation et de rotation de durées aléatoires). On notera qu'en l'absence de capacités de navigation, seul ce comportement permet de trouver des zones de recharges (sombres ou claires).
– *EvitementObstacle (EO)* : le robot effectue une marche arrière suivie, dans le cas d'un choc de côté, d'une rotation de 45° dans le sens opposé au côté du choc, et d'une rotation de 180° lors d'un choc frontal.
– *Toilette (T)* : le robot s'arrête et effectue un « toilettage virtuel », c'est-à-dire sans activation d'aucun effecteur.

Ces cinq premiers comportements consomment de l'*Energie* à un taux de 0,5 unité.s^{-1} (sur l'échelle de 0 à 255). Après normalisation, ce taux vaut donc 2.10^{-3} par seconde.

– *Repos (R)* : le robot s'arrête.

Ce comportement permet de limiter la consommation d'*Energie* : il consomme moitié moins que les autres (0,25 unité.s^{-1}, 1.10^{-3} après normalisation).

Les variations de l'*Energie Potentielle* et de l'*Energie* en fonction de la sélection de l'un ou l'autre de ces comportements durant une séquence-type de comportements sont résumées en figure 3.4. Lorsque le robot est en phase d'exploration (aussi bien que lorsqu'il évite un obstacle ou se toilette), l'*Energie Potentielle* est constante et l'*Energie* décroît. Il en va de même pour le repos, mais avec un taux de consommation de l'*Energie* moitié moindre. Lorsque le robot se recharge sur une zone sombre, il consomme toujours de l'*Energie* mais emmagasine de l'*Energie Potentielle*. Enfin, lorsqu'il se recharge sur une zone claire, il transforme ses réserves d'*Energie Potentielle* en *Energie* utilisable.

3.2.2 Implémentation et adaptations du modèle

Les paramètres internes du modèle GPR, définissant les seuils et pentes des neurones des différents modules du modèle, sont identiques à ceux utilisés par Montes-Gonzalez (2001) (tab. 3.1), si ce n'est que la modulation dopaminergique est constante (0,2) et identique dans toutes

Module	Seuil	Pente
Striatum D1	0,2	0,35
Striatum D2	0,2	0,35
NST	-0,25	0,35
GP	-0,2	1
EP/SNr	-0,2	1
Persistance	0	1
TRN	0	0,5
VL	-0,8	0,62

TAB. *3.1: Paramètres des fonctions de transfert des neurones des différents modules du modèle GPR.*

les expériences. En revanche, des modifications ont été apportées en ce qui concerne le calcul des saliences, les mises à jour du système et l'interprétation des sorties.

Modifications des calculs d'entrée

La nécessité de l'activation d'un comportement à un instant donné est évaluée par une valeur de salience, qui sert de valeur d'entrée au modèle GPR. Ces saliences sont des fonctions des variables externes (L_B, L_N, C_G et C_D), des variables internes (E_{Pot}, E et S) et du signal de persistance (P) propre à chaque comportement, provenant de la boucle de rétroaction issue du circuit thalamo-cortical (TH). Dans l'expérimentation robotique menée par Montes-Gonzalez (Montes-Gonzalez *et al.*, 2000; Montes-Gonzalez, 2001), les saliences étaient calculées par des neurones Sigma, c'est-à-dire opérant de simples sommes pondérées de leurs entrée. Dans nos expériences, compte tenu de la tâche à accomplir, deux ajouts se sont avérés nécessaires : le pré-traitement des entrées par l'utilisation de fonctions de transfert et le calcul des saliences par des neurones autorisant des multiplications entre les entrées sensorielles (neurones Sigma-Pi).

L'application de fonctions de transfert aux entrées sensorielles permet de faire dépendre la salience de certains comportements d'entrées sensorielles pré-traitées. Cela s'avère nécessaire, par exemple, pour la salience de *RechargeSurNoir* qui doit dépendre du manque d'*Energie Potentielle*, obtenu en appliquant à l'*Energie Potentielle* la fonction $f(x) = 1 - x$. Les fonctions utilisées sont répertoriées dans le tableau 3.2.

Une somme pondérée ne permet pas de faire dépendre une salience du couplage de deux variables. Or dans nos expériences, la salience de *RechargeSurNoir*, par exemple, devrait dépendre

$Rev(x)$	=	$(1-x)$
$Circ(x)$	=	$\sqrt{1-x^2}$
$f(x)$	=	x^2
$g(x)$	=	$\sqrt{x}$

TAB. *3.2: Fonctions de transfert utilisées en pré-traitement des calculs de saliences.*

du couplage de la *Noirceur* et du manque d'*Energie Potentielle*. En effet, activer *RechargeSurNoir* sur une zone claire car le manque d'*Energie Potentielle* est très fort, ou l'activer sur une zone sombre alors qu'il n'y a pas de manque d'*Energie Potentielle*, consomme de l'*Energie* sans aucun bénéfice (cf. 3.2.1.0). Nous avons donc remplacé les neurones Sigma du modèle original par des neurones Sigma-Pi permettant de multiplier entre elles des entrées du neurone avant d'effectuer la somme pondérée.

En effet, une première série de tests menée en calculant les saliences selon le mode original de calcul a mis en évidence la nécessité d'opérer ce changement. En plus de 12 heures de temps cumulé, aucune des combinaisons de poids pour les calculs de saliences testées pour le GPR comme pour le WTA n'a permis au robot de survivre plus d'une fois et demi sa durée de vie minimale (9 minutes).

Un exemple illustrant le déroulement de ces essais est présenté figure 3.5. On constate que le comportement *RechargeSurNoir* y est sélectionné deux fois de façon inappropriée pour deux raisons opposées, puis que le comportement *RechargeSurBlanc* est également sélectionné de façon inappropriée et fatale. La première activation de *RechargeSurNoir* est déclenchée par le seul manque d'*Energie Potentielle* (elle vaut 0,5 en début d'expérience), sur une zone gris neutre, ce qui consomme de l'*Energie* sans procurer d'*Energie Potentielle*. La consommation d'*Energie* augmente la salience du comportement *ExplorationAléatoire*, qui finit par prendre la main. Après une courte exploration, le robot passe sur une zone sombre (*Noirceur* aux environs de 0,5). *RechargeSurNoir* reprend donc logiquement le contrôle du robot : il est sur une source d'*Energie Potentielle* alors qu'il en manque. Ceci fait que, dans un premier temps, le robot se recharge totalement en *Energie Potentielle*. Cependant, une fois rechargé, la seule entrée sensorielle de *Noirceur* suffit à maintenir *RechargeSurNoir* actif. La consommation d'*Energie* continuant, le niveau d'*Energie* atteint de si basses valeurs que la salience du comportement *RechargeSurBlanc* vient à dépasser celle de *RechargeSurNoir*. *RechargeSurBlanc* reste alors sélectionné bien que le robot soit sur une zone sombre, aucune transformation d'*Energie Potentielle* en *Energie* n'a donc lieu, l'*Energie* continue d'être consommée jusqu'à ce qu'elle atteigne 0, mettant fin au test.

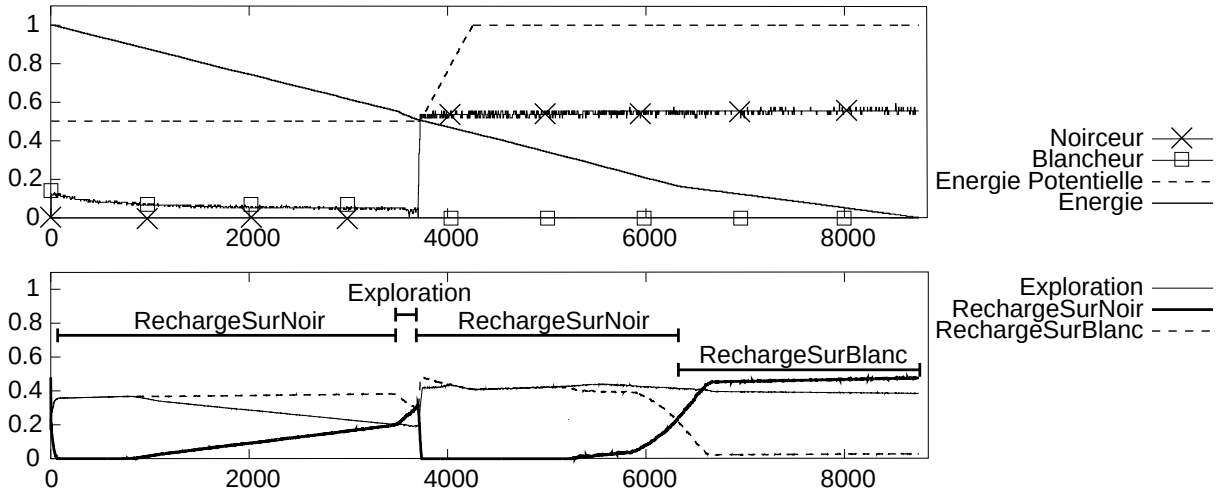

FIG. *3.5: Test typique du GPR effectué sans neurones Sigma-Pi. Haut : variables externes (*Noirceur *et* Blancheur*, contacts omis) et internes (*Energie Potentielle *et* Energie*). Bas : valeur absolue des inhibitions de sortie des comportements impliqués dans la séquence et comportement sélectionné (celui dont l'inhibition est la plus proche de 0). L'abscisse est en nombre de cycles de calcul (entre 14 et 15 par seconde).*

On constate donc dans ce test que c'est l'absence de couplage entre variables internes et externes qui est la source des sélections erronées. L'utilisation de combinaisons multiplicatives pour le calcul de salience a amélioré grandement la durée de vie du robot, quelle que soit l'architecture de sélection employée. Bien qu'étant toujours susceptible de « mourir », puisque les ressources de l'environnement ne peuvent être retrouvées que par recherche aléatoire, le robot est alors capable de survivre plusieurs heures. Toutes les expériences menées dans ce cadre expérimental utilisent donc des neurones Sigma-Pi pour le calcul des saliences.

De tels neurones ont déjà été utilisés pour résoudre des problèmes similaires, que ce soit dans le cadre de l'apprentissage dans les réseaux de neurones (Rumelhart et McClelland, 1986; Gurney, 1992) ou de celui du traitement du contexte (Balkenius et Moren, 2000). Il convient cependant de se demander si leur usage n'est qu'une solution relevant de l'ingénierie ou si une correspondance avec les données biologiques existe. Mel (1993) affirme que les arborisations dendritiques des cellules pyramidales du neocortex peuvent calculer des fonctions complexes de ce type. Il est donc raisonnable de présumer que des fonctions du second ordre sont extraites par les neurones du cortex ou du striatum qui calculent les saliences.

Le détail des calculs de saliences est donné dans le tableau 3.3.

Comportement		Calcul de la salience
ExplorationAléatoire	**WTA**	$-C_G - C_D + 0,5 \times Rev(E_{Pot}) + 0,7 \times Rev(E)$
	GPR	$-C_G - C_D + 0,8 \times Rev(E_{Pot}) + 0,9 \times Rev(E)$
EvitementObstacle	**WTA**	$3C_D + 3C_G$
	GPR	$2C_G + 2C_D + 0.5P_{EO}$
RechargeSurNoir	**WTA**	$-2L_B - C_G - C_D + 3L_N \times Rev(E_{Pot})$
	GPR	$-2L_B - C_G - C_D + 3L_N \times Rev(E_{Pot}) + 0.4P_{RSN}$
RechargeSurBlanc	**WTA**	$-2L_N - C_G - C_D + 5L_B \times Circ(Rev(E_{Pot})) \times Rev(E)$
	GPR	$-2L_N - C_G - C_D + 5L_B \times Circ(Rev(E_{Pot})) \times Rev(E) + 0.5P_{RSB}$
Repos	**WTA**	$-C_G - C_D + 0.1$
	GPR	$-C_G - C_D + 0.6P_R$

TAB. *3.3: Calcul des saliences utilisées par le GPR et le WTA dans les expérimentations.*

Mises à jour du système

Du fait de ses nombreuses boucles (dans les inhibitons mutuelles dans le striatum D1 et D2, entre NST et GPe et sur l'ensemble du modèle via la persistance), le GPR est un système dynamique. Soumis à une entrée constante, il tend à se stabiliser vers un point fixe. C'est la sortie stabilisée qui est utilisée pour déterminer la décision prise par le système, et non les valeurs intermédiaires produites lors de la période transitoire. Pour ce faire, l'implémentation robotique de Montes-Gonzalez fait autant de mises à jour du système qu'il le faut pour qu'il se stabilise avant de traiter ses inhibitions de sortie. Dans notre cas, autoriser le système à effectuer jusqu'à quatre mises à jour entre chaque nouvelle lecture des senseurs s'est avéré suffisant.

Modification des sorties

Le modèle initial génère les commandes du robot en effectuant une somme des ordres moteurs générés par chaque comportement, pondérée par leur désinhibition. Cette méthode, dite de « soft-switching », permet d'effectuer des compromis entre comportements partiellement désinhibés. Afin que les comparaisons WTA/GPR soient possibles, sachant qu'un WTA ne permet pas ce type de compromis, nous avons choisi, pour le GPR, de n'activer qu'un seul comportement à la fois, le plus désinhibé (méthode dite de « hard-switching »). En effet, nous cherchons à mettre en évidence les différences entre les modes de sélection interne du GPR et d'un WTA, et non à estimer les avantages qu'une méthode de « soft-switching » peut avoir sur une méthode de « hard-switching ».

Implémentation du WTA

Le mécanisme de sélection WTA utilisé pour les comparaisons avec le GPR est le plus simple possible. A chaque nouvelle lecture de senseurs, des saliences sont calculées à partir des mêmes outils que celles du GPR : fonctions de transfert et neurones Sigma-Pi. Les fonctions utilisées et les poids des connections sont cependant ajustés pour rendre le WTA le plus efficace possible et sont donc différents de ceux du GPR (tab. 3.3). C'est le comportement dont la salience est la plus forte qui est immédiatement sélectionné.

Détails techniques

Le contrôleur RCX du robot Lego est limité à 32 Ko de mémoire, dont une partie est utilisée par son système d'exploitation (LegOS). Du fait de cette limite et de la relative lenteur du processeur embarqué, les calculs effectués à bord du robot sont limités à la lecture et le filtrage des senseurs, à la mise à jour du métabolisme, aux comportements et aux communications. Un PC sous Linux prend en charge les calculs relatifs au modèle GPR, renvoyant au RCX les ordres résultants des informations sensorielles et métaboliques reçues. Les communications entre le RCX et le PC opèrent par l'émetteur-récepteur à infrarouge standard des Lego Mindstorms, à une fréquence d'environ 14,5 Hz.

3.2.3 Expérimentation et résultats

Expériences

Les deux architectures –WTA et GPR– ont été testées sur le même robot, confrontées aux mêmes tâches, dans le même environnement. Leurs saliences sont calculées à partir des mêmes variables –à l'exception de la persistance de chaque comportement, spécifique au GPR– et avec les mêmes outils (fonctions de transfert, neurones Sigma-Pi). Chaque système possède son propre ensemble de poids et de fonctions de transfert pour le calcul de ses saliences (cf. C.3). Ces ensembles ont été ajustés à la main durant une série d'expériences préliminaires, afin de les rendre les plus efficaces possible.

Pour chaque expérience, les deux architectures sont testées avec un nombre d'essais variable (de 5 à 10). Pour chaque essai, l'*Energie* est initialisée à 1 et l'*Energie Potentielle* à 0,5, ce qui garantit une durée de vie de 9 minutes si aucune recharge n'est effectuée. Cette durée de vie n'est dépendante que de son métabolisme virtuel, en effet ses batteries lui permettent près de 5 heures de fonctionnement continu. Un système de sélection de l'action sera considéré comme fonctionnel s'il est capable d'assurer au moins une heure de survie par essai.

Les données concernant les variables externes, internes et le comportement actif sont enregistrées à la fréquence de 14,5 Hz. Pour chaque essai, ces données sont utilisées pour calculer les mesures suivantes :

- Les médianes de l'*Energie*, de l'*Energie Potentielle* calculées à partir de toutes les valeurs enregistrées.
- La quantité moyenne d'*Energie Potentielle* extraite par seconde des ressources inépuisables de l'environnement. Cette mesure est obtenue en ne sommant que les variations positives de l'*Energie Potentielle* durant les activations de *RSN* puis en divisant le résultat par la durée totale du test.
- La médiane de la durée d'activation de chaque comportement, calculée à partir des durées de chacune des activations.
- Les fréquences d'activation de chaque comportement, calculées en divisant le nombre d'activations de chaque comportment durant l'essai par sa durée en heures.

Pour chaque expérience, les ensembles de mesures obtenus pour le GPR et le WTA sont comparées par le test U de Mann-Whitney.

L'expérience 1 compare les deux architectures en les dotant du plus petit ensemble de comportements permettant la survie (c'est-à-dire *ExplorationAléatoire*, *RechargeSurNoir*, *RechargeSurBlanc* et *EvitementObstacles*), afin d'estimer si les deux systèmes sont en mesure de résoudre cette tâche de survie et si les boucles récurrentes du GPR lui confèrent des avantages en terme de stratégie comportementale et de maintien des variables internes dans une plage de valeurs élevée.

L'expérience 2 analyse les effets de la persistance du GPR et compare plus particulièrement la capacité des deux systèmes à éviter les oscillations comportementales. Pour ce faire, le comportement de *Toilette* et la variable interne correspondante *Saleté* sont ajoutés. Les comportements de recharge et de toilettage sont mis en compétition directe, par ajustement des poids des saliences respectives, ce qui est susceptible de générer des oscillations comportementales en cas de fortes valeurs de *Saleté* durant les recharges.

L'expérience 3 compare la capacité de chacun des systèmes à économiser leur *Energie*. L'opportunité est en effet donnée à chaque système d'activer le comportement de *Repos*, en plus de ceux de l'expérience 1. Il ne permet pas au robot de bouger, et donc de trouver des zones de recharge s'il est en manque d'*Energie* ou d'*Energie Potentielle*, mais, en contrepartie, il consomme moins d'*Energie* que les autres. Il est donc opportun de l'activer lorsque les niveaux d'*Energie* ou d'*Energie Potentielle* sont suffisamment élevés et d'éviter ainsi une dépense énergétique importante.

Durée	**EA**	**RSN**	**RSB**	**EO**
GPR				
Médiane	50	253	212	34
Intervalle	46 : 52	133,5 : 356	145 : 268	31 : 38
WTA				
Médiane	46	141	139	20
Intervalle	40 : 48	98 : 246	112 : 152	20 : 20
U	24	8	3	3
	p>0,05	**p<0,01**	**p<0,01**	**p<0,01**
Fréquence	**EA**	**RSN**	**RSB**	**EO**
GPR				
Médiane	272,5	28,8	50,0	233,1
Intervalle	259,1 : 294,1	26,6 : 45,6	40,8 : 59,1	220,1 : 257,7
WTA				
Médiane	433,8	40,6	52,0	331,4
Intervalle	393,4 : 471,0	24,7 : 49,2	48,3 : 61,4	295,2 : 403,7
U	0	18	20	0
	p<0,01	**p<0,05**	**p<0,05**	**p<0,01**

TAB. *3.4: Expérience 1 : comparaison (test U de Mann-Whitney) entre les GPR et le WTA, tous essais cumulés. Haut : durées des comportements (en pas de temps). Bas : Fréquence d'activation de chaque comportement par heure.*

Expérience 1 : survie

9 essais ont été réalisés pour le GPR et 10 pour le WTA. Les deux mécanismes de sélection de l'action –GPR et WTA– se sont avérés capables de survivre bien plus d'une heure par essai, et donc de maintenir leurs variables internes dans leur *zone de viabilité*.

Ce premier résultat nous a incité à étudier plus en détail les stratégies comportementales exhibées par les deux mécanismes.

La figure 3.6 représente deux extraits des séquences comportementales du GPR (en haut) et du WTA (en bas) en cours de fonctionnement, durant un peu plus de 1min 40s. D'une part, le graphe (a) montre les saliences d'entrées du GPR, suivies, en (b), des inhibitions de sortie correspondantes et en (c) des comportements activés en conséquence. D'autre part, le graphe (d) représente les saliences d'entrée du WTA et le graphe (e) la séquence comportementale qui

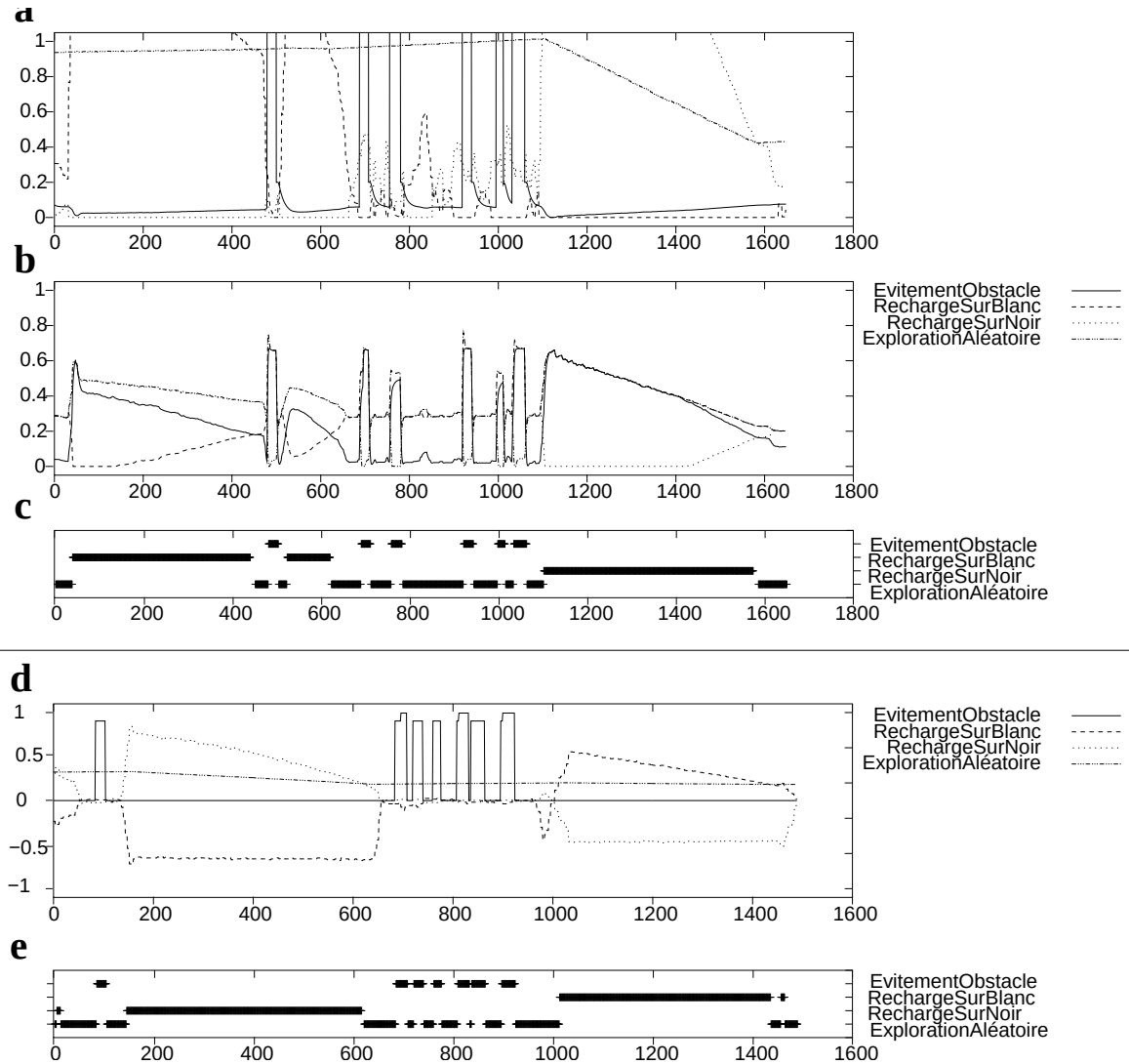

FIG. *3.6: Extraits du fonctionnement standard du GPR (a,b et c) et du WTA (d et e). GPR : (a) saliences, (b) valeur absolue des inhibitions de sortie, (c) séquence comportementale correspondante. WTA : (d) saliences, (e) séquence comportementale correspondante. L'abscisse est en nombre de cycles de calcul (entre 14 et 15 par seconde).*

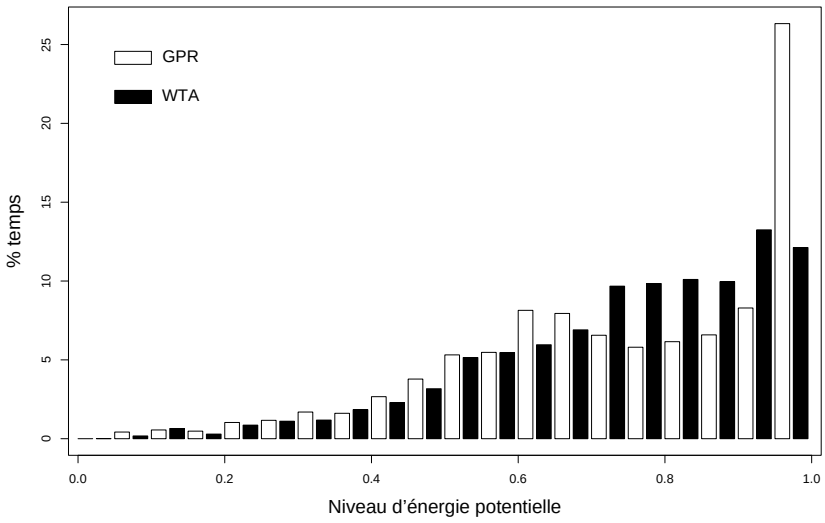

FIG. *3.7: Expérience 1 : distribution de l'*Energie Potentielle *au cours du temps, établi sur tous les essais cumulés du GPR d'une part et tous les essais cumulés du WTA d'autre part.*

en découle. On notera la différence substantielle entre les saliences d'entrée en (a) et en (d), celles du (a) étant nettement renforcée par la rétroaction positive de la persistance. Les signaux de sortie du GPR (b) montrent que l'action combinée du circuit de contrôle (*GB II*) et de la boucle de rétroaction positive (*TH*) ont amélioré le contraste des saliences d'entrée.

L'analyse des durées d'activations des comportements et des fréquences de leurs activations, résumée dans le tableau 3.4, permet de distinguer des stratégies de survie différentes pour chaque système de sélection.

En effet, le WTA compense la plus courte durée de ses comportements de recharge et d'évitement d'obstacles par des activations plus fréquentes, de sorte que les médianes d'*Energie Potentielle* et d'*Energie* des deux architectures ont finalement des valeurs similaires.

L'histogramme de la figure 3.7, qui représente le pourcentage du temps total (en ordonnée) durant lequel l'*Energie Potentielle* du robot est dans les plages de valeurs représentées en abscisse, révèle également une grande similarité des niveaux de rechargement, à l'exception d'une classe. En effet, le robot GPR maintien son *Energie Potentielle* à plus de 95% de la charge maximale pendant plus de 25% du temps. En comparaison, le WTA n'y parvient que pendant moins de 13% du temps. Cette observation suggère que le GPR se trouve dans cette *zone de confort* plus longtemps que le WTA. Pour autant, les médianes d'*Energie Potentielle* sont similaires (fig. 3.8 et tableau 3.5), cette différence n'a donc pas d'effet global significatif. Concernant l'*Energie Potentielle* moyenne extraite chaque seconde de l'environnement, elle

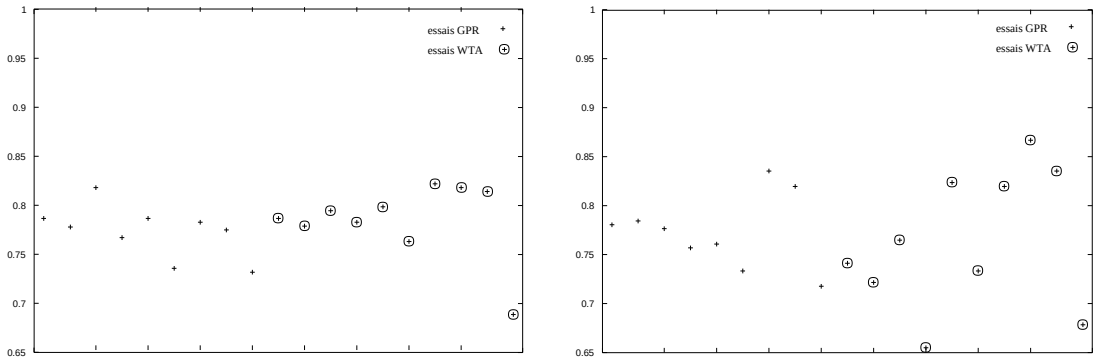

FIG. *3.8: Expérience 1 : valeurs médianes d'*Energie *(gauche) et d'*Energie Potentielle *(droite) pour chaque essai mené pour le GPR et le WTA.*

		E	**E**$_{Pot}$	**E**$_{Pot}$ **extr. (1**10^{-3}**)**
GPR	Médiane	0,78	0,75	2,3
	Intervalle	0,68 : 0,81	0,65 : 0,86	2,2 : 2,4
WTA	Médiane	0,77	0,77	2,2
	Intervalle	0,72 : 0,81	0,71 : 0,83	2,0 : 2,6
	U	26	43	27
		p>0,05	p>0,05	p>0,05

TAB. *3.5: Expérience 1 : comparaison (test U de Mann-Whitney) entre le GPR et le WTA, pour les médianes de l'*Energie, *de l'*Energie Potentielle *et de la moyenne de l'*Energie Potentielle *extraite de l'environnement chaque seconde, mesurées pour chaque essai.*

n'est naturellement pas significativement différente pour les deux systèmes, car dans cette situation tous leurs comportements ont le même taux de consommation d'*Energie*. On observe que, puisque la transformation d'*Energie Potentielle* en *Energie* est dissipative (voir equations 3.2 et 3.3), cette *Energie Potentielle* moyenne extraite chaque seconde de l'environnement ($2,2.10^{-3}$ et $2,3.10^{-3}$) est légèrement supérieure au taux de consommation d'*Energie* (2.10^{-3}) des comportements.

Cette première expérience montre que les deux systèmes sont capables d'exhiber des choix adéquats, mais qu'il utilisent des stratégies de survie différentes (recharges longues pour le GPR, recharges plus courtes mais plus fréquentes pour le WTA).

Ces différences comportementales dérivent de la durée des activations des comportements. En effet, la durée d'un comportement du robot GPR –sélectionné par le GBI– est allongée par la rétroaction positive issue de la boucle thalamo-corticale –sous partie TH du système–, sous le

contrôle du $GBII$. La valeur de persistance calculée par TH augmente la salience gagnante, favorisant donc la sélection du comportement correspondant pour les pas de temps suivant. En parallèle, les signaux inhibiteurs issus du $GBII$ et dirigés vers le NST et l'EP/SNr diminuent globalement l'activation des neurones du modèle. Sans ce contrôle, un comportement sélectionné pourrait s'auto-renforcer jusqu'à saturer le réseau et à prévenir ainsi sa désélection.

La figure 3.9 précise l'effet de la persistance sur un exemple impliquant la sélection de *RechargeSurNoir*. La salience de *RechargeSurNoir* est proportionnelle au manque d'*Energie Potentielle*, ce qui implique que l'activation de ce comportement de recharge cause la diminution de sa propre salience. Avec l'architecture WTA, une autre salience peut donc facilement interrompre ce comportement avant que l'*Energie Potentielle* ne soit totalement rechargée. Au contraire, si les fonctions de transfert et les poids des calculs de salience sont correctement ajustés, le GPR est capable se recharger plus longuement, il est en effet moins facilement interrompu, puisque la salience de *RechargeSurNoir* est renforcée par le signal de persistance.

Les expériences 2 et 3 sont destinées à analyser en quoi la persistance peut contribuer à améliorer la performance du robot dans les problèmes classiques de sélection de l'action : oscillations comportementales (expérience 2) et gestion de l'énergie (expérience 3).

Expérience 2 : oscillations comportementales

L'expérience 2 se propose de tester l'efficacité de la persistance lorsque le robot est confronté au choix de l'« âne de Buridan », où deux comportements doivent être effectués avec la même urgence.

En effet, en plus de permettre des recharges plus complètes, la persistance pourrait permettre de limiter les oscillations comportementales, c'est à dire les changements trop fréquents d'un comportement à un autre qui ne permettent pas de mener correctement ni l'un ni l'autre des comportements impliqués. Ainsi qu'il l'a été mentionné précédemment (voir 1.2.2), éviter ces oscillations lorsqu'elles sont néfastes pour la survie est un problème auquel la plupart des architectures de sélection de l'action dérivant d'une approche ingénieur sont confrontées.

Notre implémentation robotique était jusqu'à présent rarement mise en situation d'oscillations comportementales. Le robot étant incapable de voir à distance, il ne peut à la fois percevoir une zone d'« ingestion » et une zone de « digestion », donc le problème de se diriger vers une ressource ou une autre –exemple classique– ne se pose jamais. La seule situation dans laquelle il est en mesure d'osciller se présente lorsqu'il est sur une zone de « digestion » alors que son *Energie Potentielle* et son *Energie* sont basses. En effet, activer *RechargeSurBlanc* va lui redonner un peu d'*Energie*, mais va consommer de l'*Energie Potentielle*, ce qui de-

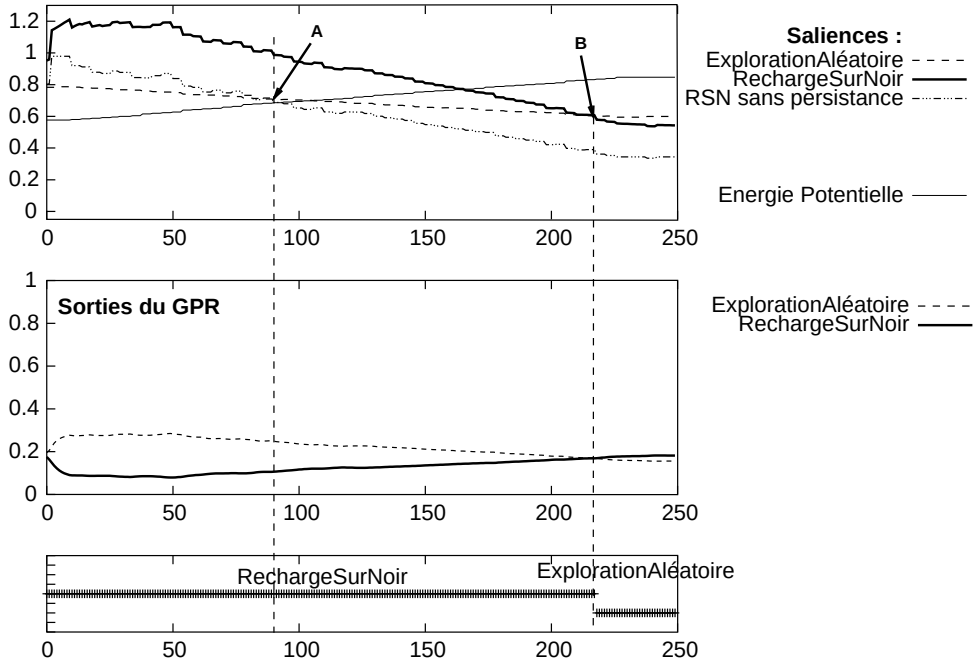

FIG. *3.9: Effet de la persistance dans le GPR. Haut : Saliences de* ExplorationAléatoire *et* RechargeSurNoir *avec et sans persistance, niveau d'*Energie Potentielle. *Milieu : valeur absolue des inhibitions de sortie des deux comportements. Bas : séquence comportementale résultante. Sans persistance, le changement de comportement aurait lieu en (A) ($E_{Pot} \simeq 0,7$), avec la persistance, il est retardé à (B) ($E_{Pot} \simeq 0,9$). L'abscisse est en nombre de cycles de calcul (entre 14 et 15 par seconde).*

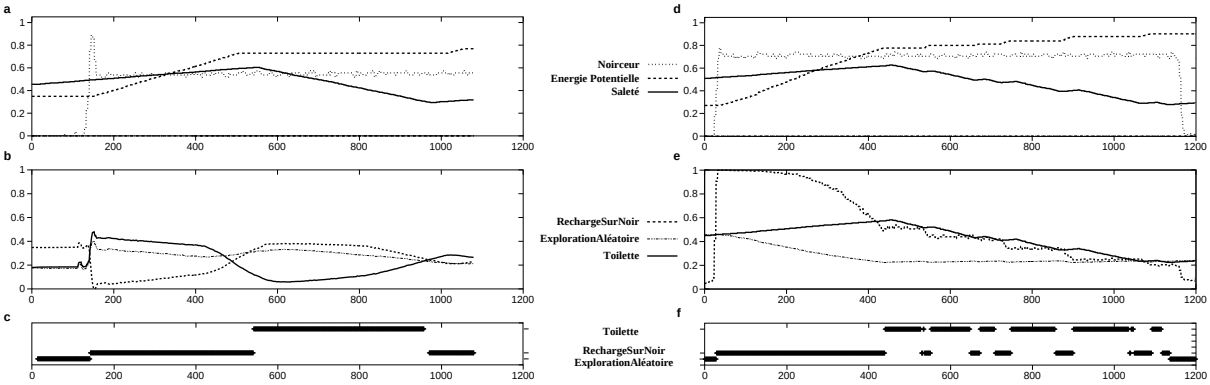

FIG. *3.10: GPR (gauche) et WTA (droite) en situation d'oscillations comportementales. (a) et (d) senseurs internes (*Energie Potentielle *et* Saleté*) et externes (*Noirceur*) déterminants dans le choix, (b) inhibitions de sortie du GPR, (e) saliences du WTA, (c) et (f) séquences comportementales correspondantes. L'abscisse est en nombre de cycles de calcul (entre 14 et 15 par seconde).*

vrait l'encourager à amorcer une *ExplorationAléatoire* pour trouver une zone d'« ingestion ». Cependant, cette *ExplorationAléatoire* va consommer de l'*Energie*, et si le robot se trouve toujours la zone de « digestion » après ce premier déplacement, il risque d'à nouveau s'arrêter pour une *RechargeSurBlanc*, et ainsi de suite. Cette situation risque de consommer les dernières réserves d'*Energie*, là où une recharge complète, non-interrompue, en *Energie*, suivie d'une phase d'*ExplorationAléatoire* assez longue pour trouver une zone d'« ingestion » serait plus efficace.

Cette situation se produit relativement rarement et ne dure que si les phases d'*Exploration-Aléatoire* ne permettent pas au robot de sortir de la zone de « digestion ». Conséquemment, pour les besoins de l'étude, une situation d'oscillations comportementales beaucoup plus nette a été mise en place par l'ajout au répertoire du robot du comportement *Toilette*. Ce comportement n'a pas d'effet particulier sur la survie du robot, si ce n'est qu'il consomme autant d'*Energie* que les autres. Il est cependant mis en compétition avec les autres comportements, puisque la salience qui lui est associée est proportionnelle à la *Saleté*, qui croit continuellement. Ainsi, dans la situation proposée, le robot est sur un carreau sombre de l'environnement (source d'*Energie Potentielle*), est en manque d'*Energie Potentielle* (entre 0,2 et 0,4) et est sale (*Saleté* aux alentours de 0,5). Il est donc susceptible d'hésiter entre l'activation de *RechargeSurNoir* et celle de *Toilette*, ainsi qu'on le constate en figure 3.10.

L'ajustement de la valeur du poids associé à la persistance dans le calcul des saliences des deux comportements (ici 0,4 pour *Toilette* et *RechargeSurNoir*) permet de régler le GPR pour qu'il limite sa propension à osciller. Il peut d'ailleurs être ajusté non seulement pour donner de l'importance à la persistance de certains comportements mais également pour autoriser les changements soudains d'autres comportements en négligeant leur persistance dans le calcul de salience. Ainsi, dans notre implémentation robotique, les comportements de recharges ont des poids associé à la persistance de 0,4 pour *RSN* et 0,5 pour *RSB*, afin qu'elles puissent être complètes, alors que d'autre part, le comportement d'exploration (*EA*), qui doit pouvoir facilement être interrompu si le robot se cogne ou passe sur une ressource, a un poids nul associé à sa persistance.

N'ayant pas de mécanisme équivalent à la persistance, le WTA est, lui, condamné à osciller de façon incontrôlable.

Cette expérience montre que le mécanisme de persistance du GPR permet d'ajuster, d'un comportement à l'autre, la facilité à être interrompu, afin d'éliminer les oscillations indésirables, et celles-là seules.

Expérience 3 : gestion de l'énergie

Malgré son absence de persistance, le robot WTA survit dans l'expérience 1 grâce à des comportements de recharge plus fréquents. Cependant, la courte durée de ces activations et le faible pourcentage du temps qu'il passe totalement rechargé suggère qu'il est susceptible de passer moins de temps que le GPR dans une *zone de confort*.

Cela n'a pourtant pas été clairement mis en évidence (voir tableau 3.5), car le GPR n'a pas été mis en situation de profiter de son avantage. En effet, dans l'expérience 1, lorsque les jauges d'*Energie Potentielle* et d'*Energie* du robot sont pleines, aucun des quatre comportements de son répertoire ne s'avére approprié : il n'a pas besoin d'explorer l'environnement à la recherche de ressources, pas plus que de se recharger ou d'éviter des obstacles. Il est pourtant forcé de sélectionner l'un d'eux et de consommer l'*Energie* correspondante.

L'ajout au répertoire du robot du comportement *Repos* –qui immobilise le robot mais consomme peu d'*Energie*– permet de comparer la capacité des deux systèmes à économiser de l'énergie (potentielle ou utilisable) lorsqu'aucun des quatre comportements de survie n'est adapté. La salience de ce comportement a été ajustée de façon à ce qu'il joue le rôle de comportement par défaut, pour le GPR autant que pour le WTA : sa salience reste toujours faible, il ne peut donc prendre la main que lorsqu'aucun autre comportement n'a besoin d'être exécuté.

5 essais ont été réalisés pour le GPR et 6 pour le WTA. Ils ont tous réussi la tâche de survie. De même que dans l'expérience 1, les deux architectures ont été capables de maintenir les variables internes du robot dans sa *zone de viabilité*.

Là encore, les statistiques des comportements ont été analysées plus en détail. Les comparaisons des durées et fréquences des comportements (tableau 3.6) montrent que les deux architectures activent *Repos* avec un fréquence similaire, mais que les durées correspondantes sont significativement plus longues pour le GPR. Conséquemment, il consomme moins d'*Energie* que le WTA et peut donc se recharger moins souvent que dans l'expérience 1. Il extrait donc moins d'*Energie Potentielle* de l'environnement que le WTA, et ce, de façon significative (tableau 3.7, dernière colonne).

Les recharges plus fréquentes du WTA lui permettent de maintenir un niveau d'*Energie* global plus élevé, mais la conjonction de conversions plus fréquentes d'*Energie Potentielle* en *Energie* et des recharges incomplètes l'empèche d'atteindre un haut niveau d'*Energie Potentielle* (fig. 3.12 et tableau 3.7).

Au contraire, le GPR atteint un très haut niveau global d'*Energie Potentielle* car il tire avantage de ses capacités, d'une part, à économiser l'énergie, et d'autre part, à se recharger plus longuement à chaque activation d'un comportement de recharge.

Durée	**EA**	**RSN**	**RSB**	**EO**	**R**
GPR					
Médiane	52,5	302	294	34	1728
Intervalle	49 : 56	233,5 : 348	233 : 308	33 : 35	1445 : 2116,5
WTA					
Médiane	48	161	150	20	485
Intervalle	48 : 49	132 : 192	120 : 168	20 : 24	340 : 568
U	1	0	0	2	0
	p<0,01	**p<0,01**	**p<0,01**	**p<0,05**	**p<0,01**
Fréquence	**EA**	**RSN**	**RSB**	**EO**	**R**
GPR					
Médiane	195,4	27,2	45,0	143,6	10,1
Intervalle	182,4 : 239,7	16,0 : 32,3	29,6 : 56,9	137,5 : 175,6	9,3 : 12,7
WTA					
Médiane	427,2	52,1	74,2	306,4	8,8
Intervalle	408,7 : 436,5	45,1 : 61,5	57,2 : 82,0	301,4 : 314,0	4,7 : 9,6
U	0	0	0	0	13
	p<0,01	**p<0,05**	**p<0,05**	**p<0,01**	p>0,05

TAB. *3.6: Expérience 3 : comparaison (test U de Mann-Whitney) entre les GPR et le WTA, tous essais cumulés. Haut : durées des comportements (en pas de temps). Bas : Fréquence d'activation de chaque comportement par heure.*

		E	**E**$_{Pot}$	**E**$_{Pot}$ **extr. (1**10^{-3}**)**
GPR	Médiane	0,78	0,94	1,8
	Intervalle	0,76 : 0,79	0,88 : 0,97	1,7 : 1,9
WTA	Médiane	0,8	0,81	2,2
	Intervalle	0,79 : 0,82	0,80 : 0,86	2,1 : 2,2
	U	1	0	0
		p<0,01	**p<0,01**	**p<0,01**

TAB. *3.7: Expérience 3 : comparaison (test U de Mann-Whitney) entre le GPR et le WTA, pour les médianes de l'*Energie*, de l'*Energie Potentielle *et de la moyenne de l'*Energie Potentielle *extraite de l'environnement chaque seconde, mesurées pour chaque essai.*

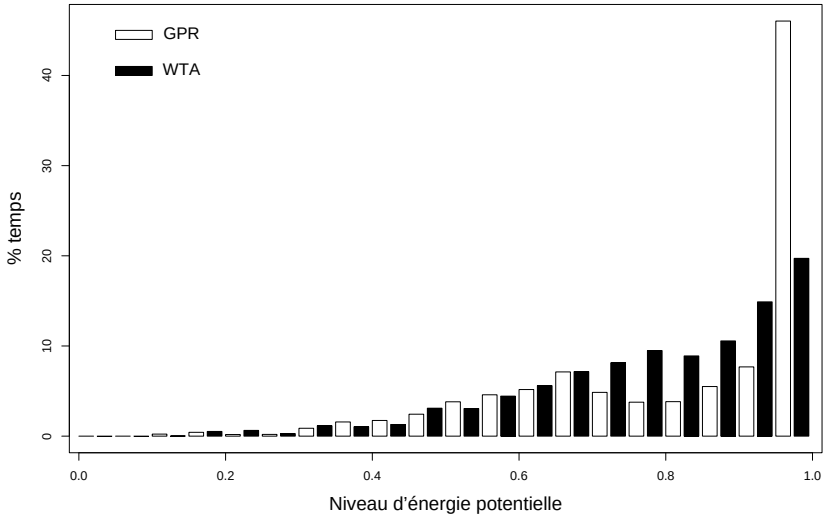

FIG. *3.11: Expérience 3 : distribution de l'*Energie Potentielle *au cours du temps, établi sur tous les essais cumulés du GPR d'une part et tous les essais cumulés du WTA d'autre part.*

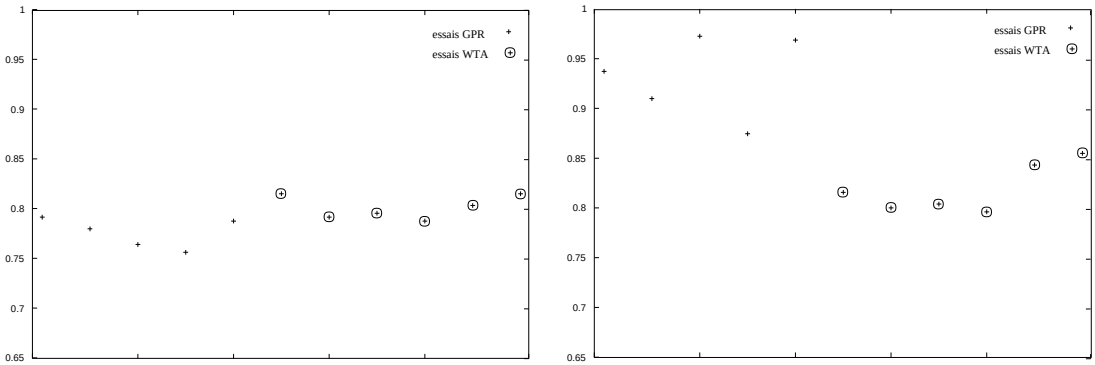

FIG. *3.12: Expérience 3 : valeurs médianes d'*Energie *(gauche) et d'*Energie Potentielle *(droite) pour chaque essai mené pour le GPR et le WTA.*

Ainsi, bien qu'il extraie moins d'*Energie Potentielle*, le GPR atteint un niveau médian d'*Energie Potentielle* plus élevé, étant à plus de 95% de recharge durant plus de 45% du temps (fig. 3.11).

Cette troisième expérience montre, d'une part, que la persistance du GPR qui lui permet d'effectuer des recharges plus longues peut avoir un effet bénéfique sur sa consommation d'énergie vis-à-vis du WTA, si tous les comportements n'ont pas le même coût.

D'autre part, il apparaît que ces recharges longues lui permettent de consacrer plus de temps à d'autres tâches que celles liées à sa survie. Il s'agit ici du *repos*, mais dans le cadre d'un robot construit non seulement pour survivre mais exécuter des tâches utiles, cette propriété semble essentielle.

Enfin, le fait que le GPR maintienne un niveau plus élevé de réserves énergétiques (son *Energie Potentielle*), le met en position de plus facilement survivre en cas de raréfaction temporaire des ressources.

3.3 Discussion

En étendant les travaux de Gurney, Prescott et Redgrave (Gurney *et al.*, 2001a; Gurney *et al.*, 2001b), notre objectif était d'explorer le rôle des ganglions de la base comme éventuel substrat de la sélection de l'action chez les vertébrés. Nous avons montré que le modèle GPR, sans l'ajout de capacités d'apprentissage, peut être réglé de sorte à générer la sélection de successions de comportements adaptée à la résolution d'une tâche de survie minimale. La robustesse du modèle semble attestée par le fait qu'il s'est avéré efficace dans deux expériences robotiques, sur deux plate-formes et pour la résolution de deux tâches différentes. La comparaison avec un WTA, capable lui aussi de résoudre la tâche, a permis de mettre en valeur des propriétés adaptatives spécifiques au modèle GPR.

Toutes ces propriétés sont issues des effets de la persistance sur la sélection. Elle maintient les réserves du robot à des niveaux plus confortables, lui permettant de mieux faire face à une éventuelle pénurie temporaire de ressources. Elle sert à éviter les oscillations comportementales lorsque cela s'avère nécessaire, l'un des principaux problèmes de sélection de l'action. Elle permet aussi au robot GPR d'économiser de l'énergie en activant plus longuement un comportement de faible coût énergétique lorsqu'aucun autre comportement n'est contextuellement pertinent.

Un dernier effet adaptatif que nous avons observé, sans que nous l'ayons testé expérimentalement, est que la persistance permet au robot d'« anticiper » les opportunités d'activation d'un comportement. Par exemple, du fait de la faible fréquence de communication entre le robot

et le PC, le robot WTA s'arrête régulièrement *après* qu'il soit passé au dessus de la partie la plus sombre (ou la plus claire) d'un carreau comportant un gradient, à cause du délai existant entre l'envoi de la lecture senseur correspondante et la réception de l'ordre. De son côté, le GPR parvient en général à s'arrêter proche du centre. Ceci s'explique par le fait que la salience du comportement de recharge en cause augmente légèrement lorsque le robot pénètre la zone sombre (ou claire). Bien que cela ne soit pas suffisant pour générer un changement du comportement sélectionné, la rétroaction positive permet d'augmenter cette salience de telle sorte que, lorsque le robot atteint le centre, il est capable de sélectionner l'action appropriée plus rapidement. Cette rapidité de réponse accrue est rendue possible par le gradient de noirceur (ou blancheur) qui sert d'indice annonciateur du changement de comportement.

L'importance de la persistance comme processus adaptatif du comportement animal a déjà été soulignée par les éthologues (McFarland, 1971a) cherchant à expliquer comment une activité pouvait continuer, alors qu'ils supposaient une baisse rapide de la pulsion la déclenchant. Ils proposèrent l'existence de mécanismes de rétroaction positive ou d'hysteresis initiés au commencement d'une période d'activité, permettant à l'animal de maintenir cette activité jusqu'à ce qu'elle soit suffisamment satisfaite (Wiepkema, 1971). Dans le modèle éthologique de Houston et Sumida (1985), par exemple, la persistance était induite lors de la compétition de deux systèmes motivationnels indépendants par un circuit de rétroaction positive similaire à une version simplifiée du modèle GPR.

Les expériences que nous avons menées fournissent une utile démonstration robotique de ce principe et de l'hypothèse de Redgrave *et al.* (1999a) selon laquelle la boucle thalamo-corticale des ganglions de la base pourrait être le substrat neural de ce circuit de rétroaction.

Pour être aussi efficace que le GPR, l'architecture WTA pourrait naturellement se voir dotée d'une boucle de rétroaction positive, similaire à la sous partie *TH* du modèle GPR, mais un système de contrôle, jouant le même rôle que le *GBII*, évitant la surcharge du système, serait alors nécessaire. On peut donc considérer que du point de vue de la robotique adaptative, l'étude de ce modèle biomimétique procure des pistes quant aux fonctionnalités à incorporer à des mécanismes de sélections de type ingénieur.

On pourrait néammoins supposer a priori une supériorité bien plus importante du GPR sur le WTA que celle qui a été mise en évidence dans ces expériences. Il est possible que certaines modifications méthodologiques puissent accentuer leurs différences.

D'une part, la prise en compte de coûts énergétiques liés aux changements de comportements, dont l'impact avait déjà été évoqué par McFarland (1989), pourrait pénaliser les transi-

tions comportementales fréquentes du WTA. Outre les conséquences énergétiques de tels coûts, des conséquences en termes mécaniques (usure prématurée et panne d'effecteurs) pourraient être extrèmement préjudiciables à un robot réel. Il serait donc judicieux de les évaluer et de les incorporer dans une prochaine implémentation.

D'autre part, la prise en compte de comportements plus nombreux aurait pu révéler l'efficacité de la sélection du GPR et en particulier de sa boucle de contrôle dans la gestion de nombreux comportements et lors de l'ajout de nouveaux comportements en cours de fonctionnement. Rappelons en effet que les ganglions de la base sont supposés être un système centralisé de sélection de l'action ayant évolué chez les vertébrés, leur ayant permis de gérer un large répertoire de comportements complexes en évitant des connections nerveuses trop nombreuses et coûteuses (Prescott, 2001). Des expériences préliminaires effectuées avec 7 comportements (ajout de comportements de remontée de gradient de teinte) ont montré la robustesse du GPR à cet égard, sans que cette caractéristique ait été systématiquement analysée. Lors de l'intégration de la navigation –décrite au chapitre 4– cette qualité sera exploitée et rediscutée.

Enfin, il faut évoquer deux limitations inhérentes à la biorobotique évoquées en introduction, liée au fait que l'architecture modélisée est supposée avoir évolué pour résoudre des problèmes environnementaux précis et qu'elle n'est qu'une sous-partie du système nerveux.

Dans notre expérience, nous avons vu qu'il était nécessaire d'analyser finement les séquences comportementales pour trouver des différences entre GPR et WTA. Il est possible que, dans notre cas, les problèmes posés au robot dans l'enceinte environnementale ne soient pas en parfaite adéquation avec la solution proposée par le GPR, qu'ils soient trop simples ou seulement indirectement liés à sa fonctionnalité.

De plus, les ganglions de la base sont en interactions avec de nombreuses autres structures nerveuses, qui ont été soit grossièrement modélisées, soit ignorées (Berthoz, 2003). Il est donc également possible que le test du GPR seul, sans modèle du cortex ou du cervelet, par exemple, ne permette pas de mettre en valeur d'éventuelles propriétés issues d'une synergie avec ces autres structures.

Malgré ces limitations, le bilan des expérimentations est positif puisque le GPR assure le maintien des variables du robot à la fois dans la *zone de survie* mais de surcroît dans la *zone de confort*. Doté d'une telle architecture de sélection de l'action, Psikharpax pourra exhiber une partie des fonctionnalités attendues de ce projet (voir 0.2.3) :

(i) utiliser ses réflexes de base pour se déplacer dans son environnement et éviter les obstacles qui s'y trouvent,

(v) utiliser un système motivationnel pour sélectionner le but courant à satisfaire,

(vii) contrôler son bilan énergétique, notamment par une alternance de périodes d'activité et de repos.

Toutefois, pour assurer les autres fonctionnalités dont Psikharpax doit être doté, ce modèle doit bénéficier de capacités de navigation et d'apprentissage.

En effet, d'une part, lorsque le robot est en manque d'*Energie Potentielle* ou d'*Energie*, il ne peut retrouver directement les ressources qu'il a déjà visité : il est incapable de se localiser dans son environnement, d'y localiser les ressources déjà découvertes et d'utiliser ces connaissances pour planifier une trajectoire permettant de rallier ces ressources connues.

D'autre part, le choix des fonctions de transfert, des variables à coupler et des poids permettant de calculer les saliences est réalisé manuellement. Ceci est possible lorsque le nombre de comportements dont les saliences doivent être ajustées les unes par rapport aux autres n'est pas trop élevé. En effet, au-delà d'une dizaine de comportements différents, il semble que cette approche soit vouée à l'échec. De plus, en cas de modification des conditions environnementales durant l'existence de l'animat, comme la raréfaction des ressources ou la diminution de leur valeur énergétique, aucun mécanisme d'adaptation ne permet à l'animat de modifier en conséquence son système de sélection.

L'intégration d'un modèle de navigation sera l'objet du prochain chapitre. La construction d'une carte, la localisation dans cette carte et son exploitation pour planifier des trajectoires ne sont pas traitées directement par les ganglions de la base, mais par l'hippocampe, en interaction avec le cortex préfrontal. Nous nous attacherons donc dans ce travail à modéliser la façon dont les ganglions de la base traitent les informations spatiales à partir d'un modèle de navigation existant.

L'intégration de processus d'apprentissage dans les ganglions de la base ne sera pas effectué dans le cadre de cette thèse. En effet, les ganglions de la base semblent être le siège d'apprentissages différenciés (stimulus-réponse ou S-R, orienté vers un but, nouveauté, etc.), plus complexes que ce que les modèles actuels (voir 2.2.1) proposent, qui à eux seuls demandent une investigation particulière et feront l'objet d'une thèse prochaine (Khamassi *et al.*, 2003).

Chapitre 4

Modèle biomimétique d'intégration de la navigation et de la sélection de l'action

A psychological consequence of this is the following :
when we analyze a mechanism, we tend to overestimate its complexity.
Braitenberg (1984)

4.1 Ganglions de la base et informations spatiales

Ainsi que cela a été évoqué en 2.1.4, le circuit des ganglions de la base à privilégier pour l'interface de la navigation et de la sélection de l'action est le circuit limbique « core ».

4.1.1 Codage de l'information spatiale dans le NAcc

Plusieurs études électrophysiologiques (Mulder *et al.*, submitted; Martin et Ono, 2000; Daw *et al.*, 2002) se sont intéressées à l'activité des neurones du NAcc « core » dans des tâches d'apprentissage spatial permettant d'atteindre une récompense. Elle confirment que le NAcc « core » est impliqué dans le codage d'une information spatiale liée à l'obtention d'une récompense.

L'une de ces expériences mesure l'activité de neurones du NAcc « core » chez des rats libres de leurs mouvement, effectuant une tâche d'apprentissage d'une séquence de déplacements dans un labyrinthe en croix, permettent d'aborder le problème de la nature de l'information codée dans les canaux du NAcc « core ».

Dans cette expérience (Albertin *et al.*, 2000), les rats sont d'abord entraînés, dans une phase préparatoire, à aller à l'extrémité d'un bras du labyrinthe boire de l'eau lorsque cette extrémité est éclairée. Une fois ce comportement acquis, la phase d'apprentissage consiste à allumer successivement les 4 extrémités, la quantité d'eau donnée en récompense étant décroissante d'un

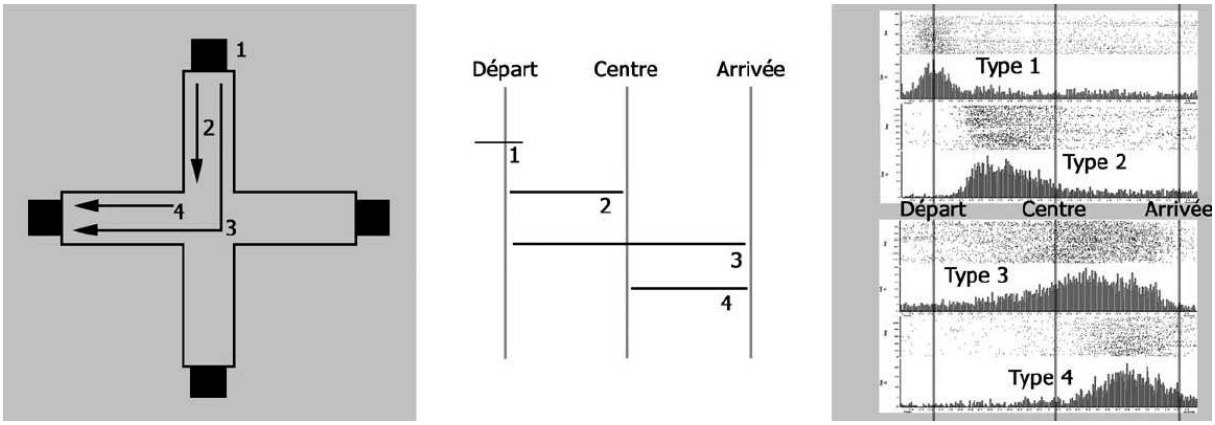

FIG. *4.1: Principaux type de cellules observés dans l'expérience du labyrinthe en croix par (Khamassi, 2003), correllés à 1) la consommation d'eau, 2) au déplacement d'un réservoir vers le centre, 3) à une séquence complète de déplacement d'un réservoir à un autre et 4) au déplacement du centre vers le réservoir.*

bras à l'autre (voir fig. 4.1). Lors de la phase de rappel, les quatre bras sont éclairés et le rat doit reproduire la séquence apprise, en allant d'abord boire au bras associé à la plus forte récompense d'eau, puis aux autres bras par ordre de quantité de récompense décroissant. Durant cette phase de rappel, les rats doivent non seulement utiliser des informations sensorimotrices, mais également des informations spatiales et des informations liées à la récompense.

A l'occasion de cette expérience, de nombreuses classes de cellules ont été identifiées dans le noyau accumbens «core» (Mulder *et al.*, submitted; Khamassi, 2003) (fig. 4.1). La première concerne les neurones actifs lors du déplacement d'un type de point de repère à un autre (d'un réservoir à un autre, d'un réservoir au centre, du centre à un réservoir). On semble donc confronté à un codage du but. La seconde classe regroupe des neurones dont l'activité est corrélée à la consommation de l'eau (en anticipation de la consommation, durant toute la consommation, à chaque goutte d'eau). La troisième regroupe les neurones ayant des décharges ponctuelles au départ d'un point de repère ou à l'arrivée auprès d'un point de repère.

Certains neurones de ces différentes classes correspondent à des préférences spatiales (neurones qui déchargent lors du déplacement du centre vers un réservoir, pour seulement trois des quatre branches du labyrinthe). D'autres, plus rares, ont leurs taux de décharge qui varient significativement entre les phases d'apprentissage et de rappel de la tâche.

Les neurones observés semblent ainsi coder une information spatiale –voire le but à atteindre pour partie d'entre-eux– ou liée à la récompense. Ce codage de l'information au niveau des neurones individuels observés, avec ses variations spatiales et liées à la récompense, est complexe et reste encore à élucider.

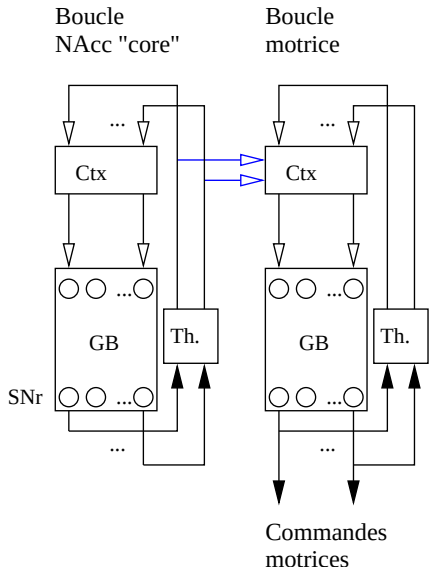

FIG. *4.2: Transmission des ordres de locomotion issu du circuit NAcc « core » vers par une boucle motrice dans le cadre d'une structuration hiérarchique des boucles. Flèches pleines : connexions inhibitrices ; flèches évidées : connexions excitatrices ; flèches en pointillés : connexions dopaminergiques.*

4.1.2 Boucle limbique « core » et ordres moteurs

La boucle limbique « core » est susceptible d'influer l'activité locomotrice du rat directement et indirectement. En effet, d'une part, le circuit du NAcc « core » pourrait la commander directement, par ses projections issues de la SNr vers des centres moteurs du tronc cérébral (Pennartz *et al.*, 1994; Groenewegen *et al.*, 1996). D'autre part, dans l'hypothèse d'une structuration hiérarchique des boucles des GB évoquée en 2.1.5 (fig. 4.2), le circuit ventral du NAcc « core » est susceptible de transmettre ses suggestions de déplacement à une boucle dorsale motrice, en charge de les mettre en œuvre.

4.2 Stratégies de navigation et ganglions de la base

Trullier (Trullier *et al.*, 1997; Trullier, 1998) établit une classification des stratégies de navigation en cinq catégories. Pour chacune d'entre elles, on cherchera à déterminer si elle est susceptible d'être traitée par les ganglions de la base, et si oui, par quelle boucle :

1. *Navigation par approche d'un objet* : simple taxie, elle ne nécessite aucune représentation interne de l'environnement et est accessible à tout animal capable de percevoir un gradient de proximité associé à l'objet à approcher. Le rat est capable d'utiliser ce type de *navigation par approche d'un objet* pour rejoindre un but perçu (par la vision ou l'odorat). Elle est susceptible d'être sélectionnée dans une boucle motrice dorsale. Cependant,

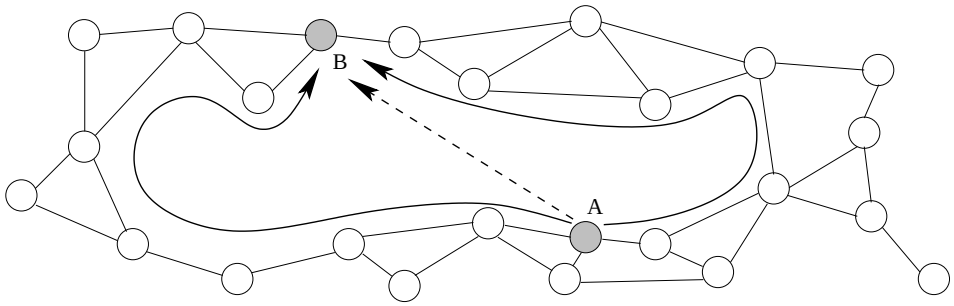

FIG. *4.3: Exemple de limitation des cartes topologiques : cartographie topologique d'un couloir, dont la partie haute et la partie basse ont été cartographiées séparément, sans que l'exploration n'ait amené jusqu'ici l'animat à passer de l'une à l'autre ailleurs qu'aux extrémités du couloir. Pour planifier un chemin permettant d'aller de A à B, seules les arêtes déjà parcourues sont utilisées, le chemin direct de A à B (en pointillé) qui est un raccourci métrique, ne peut être trouvé.*

dans des expériences de recherche de nourriture aléatoire, où seule l'approche d'objet peut être utilisée, une lésion du NAcc « core » fait baisser la performance des rats (Seamans et Phillips, 1994). Il est dont tout à fait envisageable qu'elle soit sélectionnée dans la boucle limbique « core ».

2. *Navigation par guidage* : cette approche d'un but est purement locale : la configuration d'un certain nombre de points de repère en ce but est enregistrée, de sorte qu'à proximité du but, lorsque ces mêmes points de repères sont perçus, l'animal se déplace de façon à les ajuster dans la configuration connue, ce qui le mène au but. Cette stratégie étant supposée être celle utilisée par les insectes volant pour retrouver leur nid ou des ressources précédemment visitées, elle ne sera pas considérée dans le contexte d'un rat artificiel.
3. *Navigation par action associée à la reconnaissance d'un lieu* : elle implique l'existence d'une carte cognitive et l'association à chaque lieu de la carte d'une direction de mouvement, de sorte que la reconnaissance d'un lieu déclenche le déplacement correspondant. Ce mécanisme est très proche d'un mécanisme Stimulus-Réponse (S-R), à la différence qu'ici, ce n'est pas une simple configuration de stimuli, mais une véritable reconnaissance du lieu, qui déclenche la réponse. Cette navigation est fondée sur un ensemble de règles associant configurations sensorielles et actions correspondantes. Les boucles motrices des ganglions de la base, associées à l'apprentissage et la génération de comportements « habitudes », pourraient être en mesure de mettre en œuvre cette stratégie de navigation.
4. *Navigation topologique* : dans cette approche, la carte cognitive représente des lieux et l'existence de liens entre eux permettant de passer de l'un à l'autre, indépendamment du but poursuivi. Le choix d'un déplacement permettant de rejoindre un but est alors issu

d'une planification fondée sur le graphe des liens entre lieux. Une limitation de ce type de mécanisme est que lors de la phase de planification, seuls les liens déjà existants peuvent être utilisés, une trajectoire nouvelle ne peut être générée. Ainsi, si deux lieux sont très proches physiquement alors que l'animat n'a jamais transité directement de l'un à l'autre, ils seront éloignés dans le graphe et le chemin généré afin de passer de l'un à l'autre sera en réalité un détour (fig. 4.3).

5. *Navigation métrique* : l'ajout d'informations métriques (distance et orientation entre les différents lieux) à la carte cognitive permet d'envisager de tester des raccourcis encore jamais empruntés, en générant des trajectoires entre deux lieux dont les positions relatives sont connues.

   L'existence de *cellules de lieu* dans l'hippocampe du rat, qui ne déchargent que lorsque l'animal se trouve dans des régions bien précises de l'environnement, suggère qu'il est doté d'une carte cognitive (McNaughton *et al.*, 1993) dont la nature métrique ou topologique n'est pas fermement établie. Il semble que l'hippocampe prend en charge la cartographie, l'orientation et la localisation, alors que les capacités de planification sont du ressort du cortex (prélimbique et antérieur cingulaire) (Seamans *et al.*, 1995; Floresco *et al.*, 1997). Il existe précisément des connexions excitatrices (glutamate) depuis l'hippocampe vers le cortex préfrontal, issues de l'aire CA1 et du subiculum, qui innervent les aires prélimbique et médiale orbitale du cortex préfrontal (PL/MO) ainsi que l'aire agranulaire insulaire du cortex préfrontal médian (Thierry *et al.*, 2000). Ces connexions semblent donc être en position de transmettre au système de planification les informations de localisation dont il a nécessairement besoin. Le cortex préfrontal et l'hippocampe se projettent directement sur le noyau accumbens. Globalement, les aires PL/MO du cortex préfrontal innervent le « core », et l'aire CA1 et le subiculum innervent le « shell », même si quelques projections entre PL/MO et « shell » d'une part, et CA1/subiculum et « core » d'autre part, existent également. La boucle limbique « core » serait donc en mesure de sélectionner les comportements locomoteurs issus d'une navigation *topologique* ou *métrique*.

4.3 Intégration de la navigation et de la sélection de l'action : modèles computationnels existants

De nombreux modèles computationnels de navigation inspirés de l'hippocampe ont été proposés, mais, comme cela a été évoqué dans le chapitre 1, peu nombreux sont les modèles biomimétiques qui abordent l'interfaçage entre la navigation et la sélection de l'action en traitant tout à la fois des contraintes d'un métabolisme virtuel, des interactions entre plusieurs stratégies

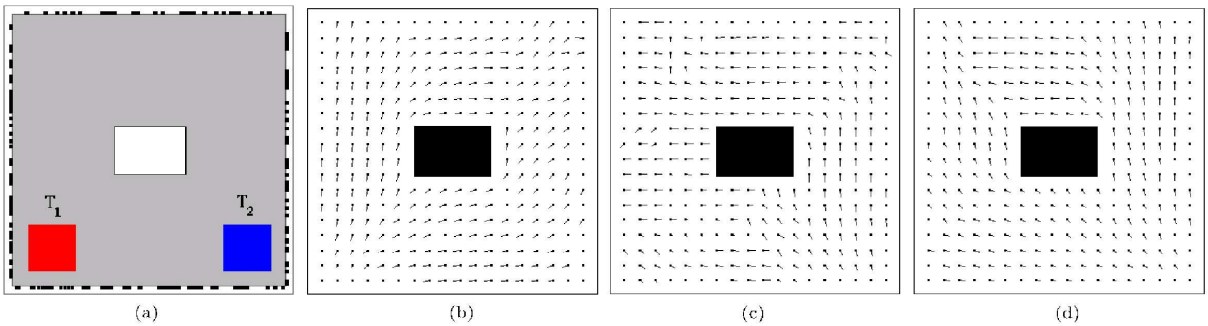

FIG. *4.4: Modèle d'Arleo : apprentissage d'une stratégie de navigation de type* action associée à la reconnaissance d'un lieu *pour deux types de buts différents (T_1 et T_2). (a) Environnement de test. Le rectangle blanc est un obstacle. (b) Les directions apprises en chaque lieu de la carte pour rejoindre le but T_1. (c) Les directions apprises de façon latente pour rejoindre T_2 pendant la recherche de T_1. (d) Les directions apprises pour rejoindre T_2. Repris de (Arleo et Gerstner, 2000).*

de navigation et du rôle des ganglions de la base dans cet interfaçage, le tout éventuellement embarqué à bord d'un robot.

4.3.1 Arleo

Le modèle proposé par (Arleo, 2000; Arleo et Gerstner, 2000) est essentiellement un modèle de navigation biomimétique, cependant il intègre quelques considérations issues de la sélection de l'action. Il utilise une fusion d'informations idiothétiques (les propres déplacements du robot) et allothétiques (les positions d'un ensemble de points de repères dans une vue panoramique) pour construire un ensemble de cellules de lieu. Il est donc capable de cartographier un environnement et de s'y localiser. Ces facultés on été testées sur un robot Khepera (©K-Team).

Le noyau accumbens est la sortie motrice du modèle, les ganglions de la base ne sont donc pas modélisés dans leur ensemble. Il est modélisé par une couche de 4 neurones, chacun codant pour une direction (Nord, Sud, Est et Ouest). La direction choisie est fonction de l'activité respective de chacun de ces neurones (méthode de « soft-switching »).

Une stratégie de navigation de type *action associée à la reconnaissance d'un lieu* est implémentée : chaque cellule de lieu est associée, par apprentissage par renforcement, aux neurones de direction du NAcc codant la direction à prendre pour rejoindre deux types de buts. Cette notion de buts multiples est prise en compte très simplement, puisque pour chaque cellule de lieu, l'apprentissage est dédoublé : deux directions sont apprises simultanément, pour rejoindre chacun des deux types de buts (fig. 4.4). Cependant, le robot n'est pas doté d'un métabolisme virtuel, il est uniquement capable de rejoindre le type de but spécifié à un instant donné par l'expérimentateur.

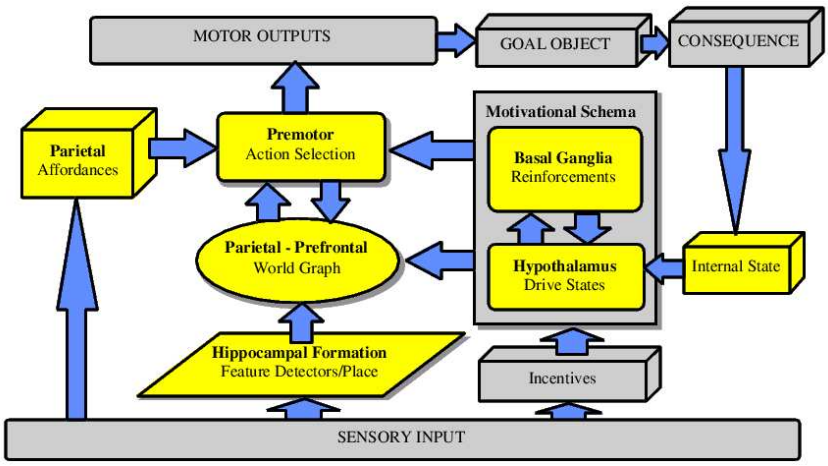

FIG. *4.5: Schéma du modèle de Guazzelli* et al.. *La structuration de l'ensemble des modules est assimilée à un certain nombre de structures nerveuses, en revanche, l'implémentation des mécanismes internes à chaque module n'est pas biomimétique. Repris de (Guazzelli* et al., *1998).*

4.3.2 Guazzelli *et al.*

Le modèle de (Guazzelli *et al.*, 1998) intègre deux stratégies de navigation (par *approche d'objets* et par *action associée à la reconnaissance d'un lieu*), implique les ganglions de la base et est doté, lui, d'un système motivationnel. Ce système n'a été testé qu'en simulation, dans des labyrinthes reproduisant des expériences comportementales chez le rat. Le rat simulé est doté d'une perception omnidirectionnelle de la distance des obstacles et des éléments attirants ou aversifs de l'environnement.

Ce modèle est issu de l'intégration de deux modèles, le Taxon-Affordances Model (TAM) –assimilable à de l'*approche d'objets*- - et le World Graph Model (WG) –implémentant une navigation par *action associée à la reconnaissance d'un lieu*. TAM modélise des taxons, qui sont des actions possibles associées aux perceptions immédiates de l'environnement (comparables aux « affordances » de Gibson (1966). Ainsi, les actions possibles associées par défaut avec les objets de l'environnement (un couloir indique une direction de déplacement, un objet coloré est susceptible d'être consommé) sont associées aux motivations par apprentissage par renforcement (un couloir ou un objet coloré particulier pourra être associé à la nourriture, par exemple). WG est une implémentation de la World Graph Theory proposée par Arbib and Lieblich (1977). Sa modélisation est fondée sur la construction d'un graphe représentant les relations topologiques entre les différents lieux visités, émulant les fonctions des cellules de lieu et de directions de la tête chez le rat. Un processus d'apprentissage par renforcement permet d'associer à chacun de ces lieux une direction à suivre pour rejoindre un but.

Plusieurs motivations sont présentes dans le modèle, elles sont supposées être calculées dans

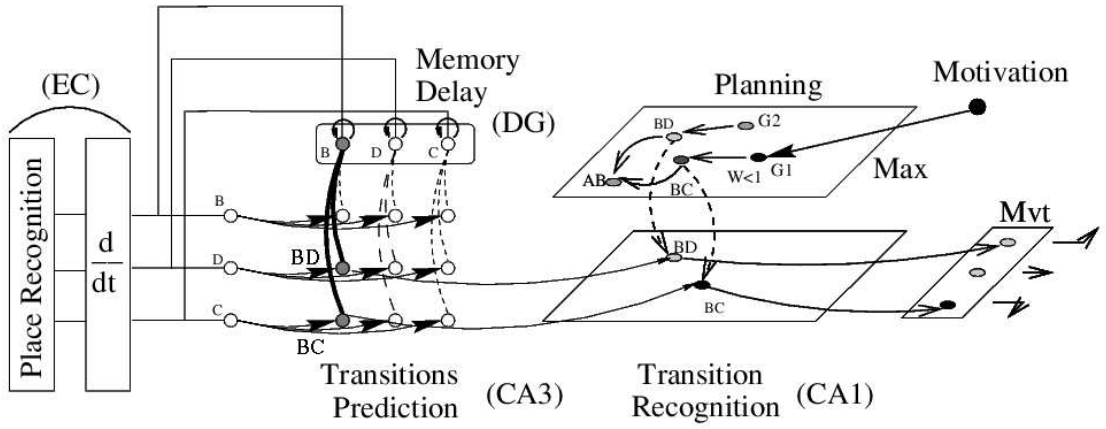

FIG. *4.6: Schéma du modèle de Gaussier* et al.. *La planification (*Planning*) est localisée dans le cortex préfrontal, la génération des ordres moteurs (*Mvt*) dans le noyau accumbens. L'activation d'une motivation propage un gradient d'activité dans la couche de planification, puis dans les transitions possibles depuis la position courante (ici BC et BD) et enfin dans les neurones du NAcc associés au mouvement correspondant à ces transitions. Un mécanisme « winner-takes-all » permet alors de sélectionner, parmi ces mouvements, celui qui est le plus activé. Repris de (Gaussier* et al.*, 2000).*

l'hypothalamus. Le niveau d'une motivation varie avec l'obtention ou au contraire le manque chronique du type de ressource correspondant (ingestion d'eau pour la motivation de soif, par exemple). L'intégration des motivations dans les processus d'apprentissage, et donc indirectement dans les choix effectués, consiste en la génération d'un signal de renforcement conditionné par la motivation prédominante. Ainsi, lorsque l'animal a faim, seule l'ingestion de nourriture génèrera un signal de renforcement positif. Le rôle des ganglions de la base est ici limité précisément à ce calcul des signaux de renforcement en fonction de l'état motivationnel, la sélection de l'action proprement-dite ayant lieu dans le cortex prémoteur (fig. 4.5). Elle somme les suggestions de déplacement issues des deux stratégies de navigation pour chacune des directions de déplacement, puis choisit la direction finale sur la base d'un mécanisme « winner-takes-all ».

On notera que malgré la présence d'un système motivationnel modulant le comportement, la survie de l'animat n'est pas contrainte par un métabolisme virtuel. Les expériences menées concernent en effet la reproduction d'expériences comportementales.

Enfin, ce modèle se place à un haut niveau de modélisation considérant les interactions entre systèmes et n'adoptant pas une démarche biomimétique dans la mise en œuvre du fonctionnement interne de chaque module considéré.

4.3.3 Gaussier *et al.*

Les travaux de (Gaussier *et al.*, 2000) sont les plus proches des objectifs de nos travaux. Ils intègrent en effet une navigation *topologique* fondée sur un modèle biomimétique de l'hippocampe et du cortex préfrontal, ainsi qu'un métabolisme artificiel contraignant la survie de l'animat.

Le système de navigation *topologique* est fondé sur une modélisation de la formation hippocampique. Les « cellules de lieu » y sont construites à partir d'informations allothétiques comparables à celles d'Arleo : la direction d'un ensemble de points de repère dans une vue panoramique de l'environnement. Ces « cellules de lieu » ont la particularité de représenter en réalité des « transitions entre lieux ». Une copie de ces cellules dans le cortex préfrontal permet la constitution d'un graphe de transitions dans lequel auront lieu les opérations de planification (fig. 4.6). Le choix d'un graphe de transitions à un graphe de lieux se justifie d'un point de vue biomimétique : si on assimile les nœuds du graphe à des neurones, dans le cadre d'un graphe de lieux, l'information sur la direction à prendre pour aller d'un lieu à un autre n'est matérialisée que par une connexion synaptique entre neurones et semble difficile à extraire. Dans un graphe de transitions, au contraire, l'activation d'un neurone correspond bien à une direction précise.

Les « cellules de transitions entre lieux » sont susceptibles d'être associées à la satisfaction d'une motivation, dans la copie corticale de la carte, par apprentissage hebbien. Le graphe des transitions obtenu autorise la mise en place d'un algorithme de planification de chemin fondé sur la diffusion dans le graphe d'activation émise par les cellules de motivation. Les transitions possibles depuis la position actuelle se voient excitées par cette activation et la transmettent aux mouvements correspondant, dans le noyau accumbens. Un simple « winner-takes-all » sélectionne alors le mouvement le plus activé, c'est à dire celui correspondant à la transition la plus proche de la ressource recherchée. A l'instar du modèle d'Arleo, le noyau accumbens est donc la couche de sortie motrice du système et les ganglions de la base ne sont pas modélisés. Le codage des mouvements dans le noyau accumbens est égocentré : trois neurones codent respectivement les actions tourner à gauche, à droite ou aller tout droit.

L'animat est doté d'un métabolisme virtuel : plusieurs motivations (faim, soif et fatigue) doivent être régulièrement satisfaites afin d'assurer sa survie, ce qui, comme nous l'avons expliqué ci-dessus, fixe implicitement les buts à atteindre lors des calculs de planification, en fonction des besoins internes (nourriture, eau ou nid), sans intervention d'un expérimentateur externe.

Ce modèle a été testé à la fois en simulation pour des expériences de survie et sur un robot réel (Koala – ©K-Team) pour des expériences plus ponctuelles mettant en évidence les capacités de planification en fonction de la motivation principale.

Enfin, la navigation topologique est la seule stratégie de navigation implémentée, elle n'est pas mise en compétition ou en collaboration avec, par exemple, une approche d'objets.

4.3.4 Bilan

En ce qui concerne la modélisation biomimétique à un niveau de réductionnisme intermédiaire, fondée sur les réseaux de neurones, elle n'est adoptée que par Arleo et Gaussier *et al.*.

En ce qui concerne la modélisation des ganglions de la base, les modèles d'Arleo et de Gaussier *et al.*, qui modélisent le fonctionnement de l'hippocampe, ne considèrent le noyau accumbens que comme une couche de sortie ayant directement accès aux ordres locomoteurs. Une sélection du mouvement final y a bien lieu –par des mécanismes de « hard » ou de « soft switching »– mais elle est simplifiée et ne tient pas compte des circuits des ganglions de la base situés en aval. Le modèle de Guazzelli *et al.* considère, lui, que cette sélection –modélisée par un « winner-takes-all »– a lieu dans le cortex prémoteur, et que les ganglions de la base ne servent qu'à générer un signal d'apprentissage modulé par les motivations.

En ce qui concerne les stratégies de navigation, les modèles d'Arleo et de Gaussier *et al.* n'en utilisent chacun qu'une seule –respectivement la navigation par *action associée à la reconnaissance d'un lieu* et la navigation *topologique*. Seul le modèle de Guazzelli *et al.* en intègre deux –l'*approche d'objets* et la navigation par *action associée à la reconnaissance d'un lieu*– dont les suggestions de déplacement sont fusionnées par une somme avant le processus de sélection.

Enfin, en ce qui concerne la sélection de l'action, seul le modèle de Gaussier *et al.* met en place un réel métabolisme virtuel contraignant l'animat à agir en temps limité pour assurer sa survie.

Il s'avère donc qu'aucun de ces modèles ne résoud à la fois des problèmes de motivation et de sélection de l'action tout en adoptant une approche biomimétique et en fusionnant plusieurs stratégies de navigation.

4.4 Modélisation de l'interface navigation/sélection de l'action

La modélisation mise en œuvre ici s'intéresse à l'interface entre les structures permettant la constitution d'une carte cognitive et son exploitation pour planifier des déplacements d'une part, et celles en charge de la sélection de l'action d'autre part (Girard *et al.*, 2003b). Le modèle chargé des tâches de navigation peut être quelconque, pourvu qu'il soit capable d'enregistrer la position des ressources rencontrées dans l'environnement et de fournir à la demande la direction à suivre pour rejoindre un type de ressource donné. Le système de navigation choisi pour notre implémentation sera décrit en 4.5.

4.4.1 Intégration de deux stratégies de navigation

Compte tenu des constatations faites en 4.2, nous nous sommes concentrés sur l'utilisation de deux types de navigations observées chez le rat : une simple *navigation par approche d'objet* et l'utilisation de la planification des déplacements à partir d'une carte cognitive de l'environnement, sans préférence particulière pour une approche *métrique* ou *topologique*. Nous avons écarté la stratégie de navigation par *action associée à la reconnaissance d'un lieu* car elle nécessite un apprentissage au sein de la boucle dorsale qui sort du cadre de notre travail.

La mise en place d'une stratégie de navigation *métrique* ou *topologique* implique de modéliser la boucle limbique (ventrale) « core » dans un rôle locomoteur. Ainsi que cela a été signalé en 4.2, l'*approche d'objets* est susceptible d'intervenir dans une boucle locomotrice ventrale ou dorsale. Nous avons opté pour la prise en charge à la fois de la navigation par *approche d'objets* et de la navigation *topologique* par la seule boucle limbique « core », considérant l'hypothèse de son contrôle direct sur la locomotion (voir 4.1.2). La boucle dorsale prendra en charge uniquement les actions motrices ne relevant pas de la locomotion, qui correspondent dans notre implémentation aux opérations de recharge.

Les fortes similarités, tant au niveau anatomique que physiologique, entre les ciruits dorsaux des ganglions de la base et celui issu du NAcc « core » (voir 2.1.2) nous incitent à modéliser chacune des deux boucles ventrale et dorsale par un GPR.

4.4.2 Sémantique des canaux

La sémantique des canaux dans le GPR, pour le circuit ventral, est modifiée : là où le GPR des circuits dorsaux utilisé au chapitre précédent effectue sa sélection parmi des comportements, nous proposons que le circuit ventral issu du NAcc « core » sélectionne directement la direction du déplacement. Ce choix est similaire à ceux faits dans les modèles computationnels de l'hippocampe ayant le NAcc comme structure de sortie, évoqués en 4.3.

Dans notre modèle, le circuit ventral issu du NAcc « core » a été doté de 36 canaux représentant des populations de neurones codant des directions allocentriques de déplacement espacées de 10°.

Il s'agit d'une simplification si l'on considère la complexité du codage de l'information dans le NAcc « core » (voir 4.1.1). Rappelons cependant que les neurones artificiels intégrateurs à fuite des canaux du modèle GPR représentent des groupes de neurones (l'ensemble des neurones présents dans un même matrisome), alors que les mesures électrophysiologiques concernent des neurones réels isolés, pouvant participer à un codage par population qui nous est pour l'instant inaccessible.

Cette nouvelle sémantique des canaux modifie la notion de « comportements » telle qu'elle est présentée au chapitre 3. En effet, un canal est maintenant associé à une direction de dépla-

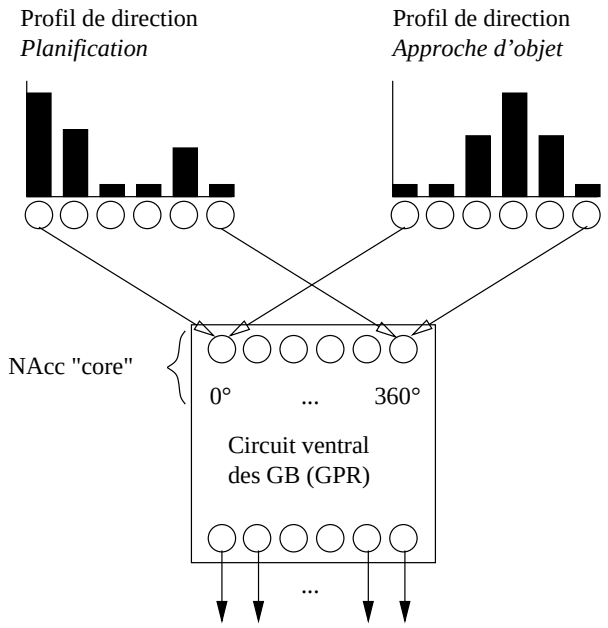

FIG. *4.7: Les profils de direction sont les vecteurs d'entrée du circuit ventral « core ». Sur ce schéma, seuls deux profils et 6 directions par profil (au lieu de 36) sont représentés.*

cement, de sorte que plusieurs comportements liés à la navigation (comme, par exemple, l'approche d'objet et la navigation topologique) peuvent ici contribuer à la salience associée à une seule direction. Cela nécessite que chaque suggestion de déplacement, émanant, par exemple, du système de planification ou d'approche d'objet, soit transmise au NAcc « core » sous la forme d'un vecteur de 36 valeurs représentant le degré d'incitation à se déplacer dans chacune des 36 directions. Ces vecteurs de 36 valeurs seront désignés sous le vocable de « profils de direction » (voir fig. 4.7) et seront représentés suivant la convention qui veut que les vecteurs soient en gras.

Du point de vue de la résolution de problèmes de sélection de l'action, cette capacité à combiner les profils de direction émanant de plusieurs stratégies de navigation avant de faire un choix de direction, plutôt que de choisir une navigation et de ne se fier qu'à sa suggestion de déplacement, permet d'effectuer des compromis similaires à ceux réalisés par le système de Rosenblatt et Payton (voir 1.2.2). La combinaison de suggestions de déplacement issues de la planification et d'autres modules, tels que l'approche d'objet ou, comme on le verra plus tard, de l'exploration ou du retour en zone connue est conforme à l'idée des « plans-as-resources », proposée par Agre et Chapman (1990). Les résultats des opérations de planification ne sont pas en effet considérés comme des ordres absolus, mais comme des suggestions à considérer parmi d'autres suggestions.

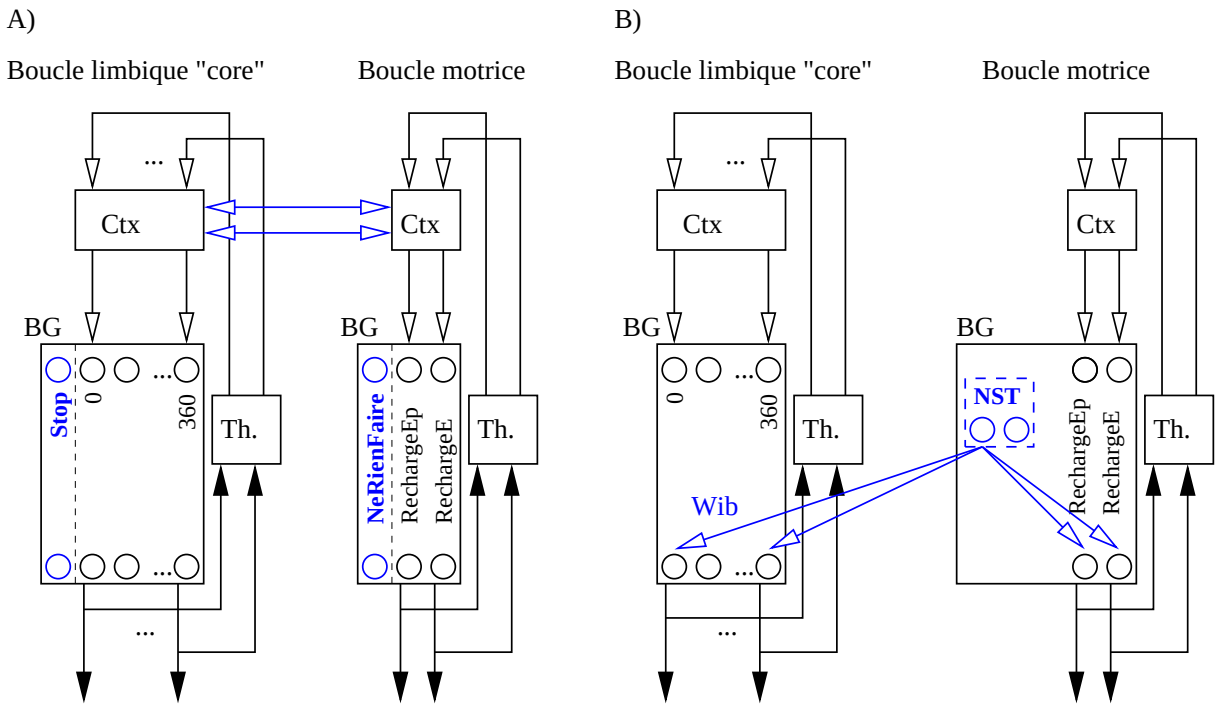

FIG. *4.8: Hypothèses d'interconnexions entre boucles des ganglions de la base modélisées. A) voie cortico-corticale : ajout de canaux* Stop *et* NeRienFaire, *dont les saliences sont calculées grâce aux données provenant des zones corticales de la boucle voisine. B) voie trans-subthalamique : le NST de la boucle dorsale est en mesure d'exciter la couche de sortie de la boucle ventrale (excitation pondérée par W_{ib} sur l'ensemble des canaux) et donc d'empêcher la sélection d'une quelconque direction. Flèches pleines : connexions inhibitrices ; flèches évidées : connexions excitatrices.*

4.4.3 Interconnexion des deux boucles

L'interconnexion de la boucle ventrale locomotrice et de la boucle dorsale motrice (hors locomotion) doit gérer la synchronisation entre la sélection des déplacements et la sélection des comportements de recharge. En effet, lorsque ces derniers sont sélectionnés, ils doivent pouvoir influer sur la boucle locomotrice pour faire cesser tout déplacement, puisqu'ils nécessitent d'être à l'arrêt pour être efficaces.

Parmi les différentes possibilités d'interconnexions entre boucles (voir 2.1.5) permettant la résolution de ce problème, nous avons tout d'abord écarté les deux voies hiérarchiques, puisqu'elles ne permettent qu'une communication depuis une boucle ventrale vers une boucle dorsale. Les deux hypothèses restantes d'interconnexion des boucles au niveau du système cortex-ganglions de la base-thalamus ont été implémentées.

Ainsi, dans une première version du modèle, nous avons proposé l'existence, d'une part, d'un

canal *Stop* dans le circuit ventral codant l'absence de déplacement et, d'autre part, d'un canal *NeRienFaire* dans le circuit dorsal codant l'expression d'aucun comportement nécessitant l'arrêt (fig. 4.8, A). En ce cas, les interconnexions corticales permettent de calculer la salience de *Stop* à partir des mêmes données sensorielles qui déclenchent les comportements de recharge immobile, et celle de *NeRienFaire* à partir des données de navigation générant des déplacements. Cependant, il s'est avéré que, dans cette configuration, il est extrêmement difficile d'ajuster les calculs de salience des deux boucles modélisées afin d'avoir une bonne coordination des sélections. Dans les faits, soit la boucle dorsale active un comportement de recharge alors que la boucle ventrale continue de sélectionner des déplacements, soit, à l'inverse, la boucle ventrale sélectionne *Stop* alors que la boucle dorsale sélectionne *NeRienFaire*. Il est apparu à cette occasion que la modification du calcul des saliences d'une boucle nécessitait l'ajustement immédiat de celui de l'autre boucle pour conserver un semblant de coordination. Cette observation ne permet d'envisager que difficilement la mise en place d'un apprentissage sur ces connexions et ne plaide pas en faveur de cette méthode de coordination des boucles des ganglions de la base.

La seconde version du modèle propose, elle, d'utiliser la voie trans-subthalamique de la boucle motrice vers la boucle limbique « core » (fig. 4.8, B). Dans cette configuration, les excitations issues de la partie du NST dédiée à la boucle dorsale sont pondérées par un poids de connexion inter-boucle W_{ib} les atténuant et transmises aux neurones du module SNr de la boucle ventrale. L'atténuation de la transmission traduit le faible nombre de ces connexions et permet de ne pas saturer le module SNr de la boucle ventrale, qui ne serait autrement plus capable de désinhiber une quelconque direction de déplacement. Dans ce cas, la décision de s'arrêter devient implicite : lorsqu'aucune direction n'est désinhibée au delà d'un certain seuil S_{inhib}, on considère que l'animat est immobile. La sélection d'un comportement de recharge en boucle dorsale envoie un surcroît d'excitation sur les neurones de sortie de la boucle ventrale, inhibant directement les déplacements. Cette seconde proposition s'est avérée très simple d'usage, la coordination étant obtenue très facilement et le calcul des saliences d'une boucle pouvant être modifié sans devoir ajuster celui de l'autre boucle en conséquence.

4.4.4 Effets du changement de sémantique des canaux

Un certain nombre de différences mineures on été ajoutées en regard du modèle utilisé au chapitre 3. Elles concernent le calcul des saliences de la boucle ventrale, nécessairement affectés par le changement de sémantique des canaux (voir 4.4.2), les inhibitions latérales dans le striatum (voir 3.1.2) et le choix des ordres par la méthode dit de « hard-switching » (voir 3.2.2).

Les canaux de la boucle limbique « core » représentant dans notre modèle des directions de déplacement et non des comportements distincts, le calcul de leurs saliences diffèrent. En

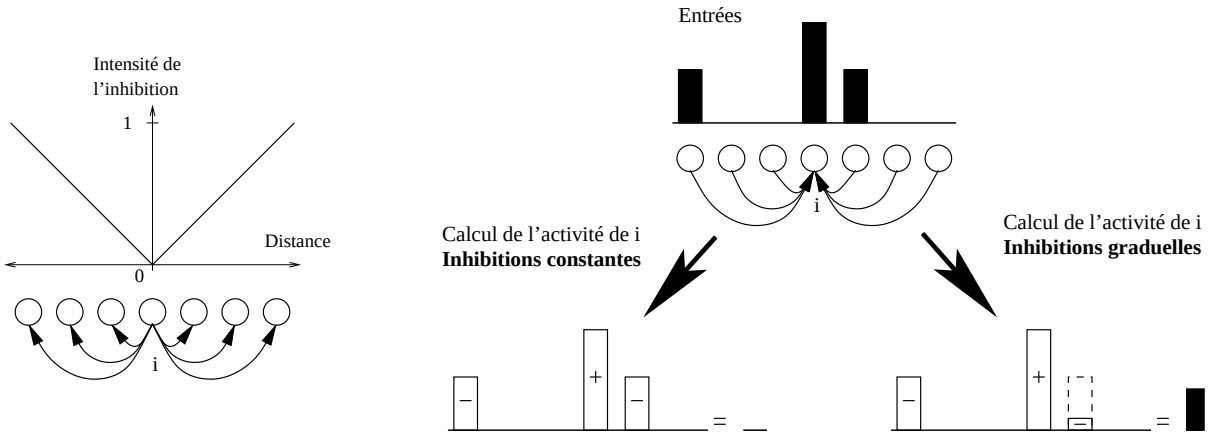

FIG. *4.9: Gauche : Inhibitions latérales graduelles linéaires utilisées dans la modélisation du NAcc « core ». Droite : Comparaison entre le mode de calcul standard des inhibitions latérales du GPR (poids constant de 1) et les inhibitions graduelles. Les inhibitions graduelles permettent à deux directions proches de ne pas se nuire entre elles.*

effet, les saliences des directions sont ici toutes calculées à partir des variables internes et des profils de directions fournis avec une unique formule. Les différences proviennent du fait que d'une direction à l'autre, un même profil ne fournit pas les mêmes valeurs. Par exemple, dans un cadre simplifié (fig. 4.7), supposons que les déplacements de l'animat ne dépendent que de deux profils de direction, celui issu de la planification et celui dépendant de l'approche d'objet (respectivement $\mathbf{Plan}$ et $\mathbf{ApprObj}$), et que le calcul de salience ne nécessite que l'usage de simples sommes pondérées, alors le calcul de la salience d'une direction i s'effectuerait comme suit :

$$Sal(i) = W_{Plan} \times \mathbf{Planif}_i + W_{ApprObj} \times \mathbf{ApprObj}_i \tag{4.1}$$

Seuls deux poids *par profi l de direction* (W_{Planif} et $W_{ApprObj}$) sont alors nécessaires, et non deux poids *par direction*.

La modification de la sémantique des canaux a également un effet sur la configuration des inhibitions latérales dans les deux sous-parties D1 et D2 du striatum. Dans le modèle inital, les comportements en compétition sont exclusifs les uns des autres, chaque comportement inhibe donc tous ses voisins avec la même intensité. Dans le cas de directions en compétition, dont les saliences sont calculées à partir de profils de direction ayant tendance à favoriser plusieurs directions contiguës, l'utilisation d'inhibitions uniformes nuit à la qualité de la sélection. En effet, deux directions contiguës de fortes saliences vont se nuire mutuellement, écrasant le contraste entre directions gagnantes et perdantes. Par exemple, lors de la présence d'une ressource occupant plus de 10° du champ visuel, le profil de direction correspondant à l'approche d'une

ressource visible va fortement activer plusieurs directions contiguës. Ces activations auront la même origine, représenteront l'approche d'un même objet et seront pour autant en compétition, nuisant ainsi à l'approche de la ressource. Dans le cas de l'utilisation d'inhibitions graduelles, inhibant faiblement les directions voisines et plus fortement celles qui sont diamétralement opposées, le contraste entre directions gagnantes et perdantes est bien meilleur et les directions voisines, activées pour une même raison, ne se nuisent plus mutuellement (fig. 4.9). Nous avons utilisé une simple inhibition croissant linéairement avec la distance entre directions. Ce qui donne, pour le calcul de l'inhibition de la direction i sur la direction j :

$$InhibGraduelle(i,j) = \frac{|i-j| \text{ mod } 18}{18} \tag{4.2}$$

Pour finir, l'exclusivité des comportements dans l'implémentation du chapitre 3 justifiait l'emploi d'une sélection de type « hard-switching » en sortie du modèle, qui n'était pas celle du modèle initial de Gurney *et al.*. Elle a été conservée dans la boucle dorsale de ce nouveau modèle. Le choix de direction de déplacement effectué en boucle ventrale est tout à fait compatible avec un mécanisme de « soft-switching », où la direction choisie serait la résultante du degré de désinhibition des différentes directions. La direction Dir choisie par la boucle ventrale est la direction du vecteur résultant de la somme de 36 vecteurs, chacun correspondant à une direction et de module égal au degré de désinhibition de cette direction en sortie du GPR, en tenant compte du seuil S_{inhib} au delà duquel aucun mouvement n'est généré :

$$desinhib(i) = \begin{cases} (S_{inhib} - SortieGPR(i))/S_{inhib} & \text{si } inhib(i) > S_{inhib} \\ 0 & \text{sinon.} \end{cases} \tag{4.3}$$

$$Dir = Arg(\sum_{k=1}^{36} desinhib(k).e^{i.(10\times k\times \frac{180}{\pi})}) \tag{4.4}$$

4.5 Système de navigation

4.5.1 Choix

Comme l'objet de notre travail n'est pas de modéliser l'hippocampe et le cortex préfrontal, ainsi que nous l'avons déjà précisé, mais de modéliser l'interface entre les ganglions de la base et ces structures, le choix du système de navigation s'est fait en fonction de critères de fonctionnalités et non de biomimétisme. Placé dans un environnement nouveau, l'animat devait être capable, en temps réel, de construire une carte, de s'y localiser, d'y enregistrer la localisation d'éventuelles ressources, de fournir une direction permettant de rejoindre une ressource de type

donné à la demande et d'être suffisamment robuste pour pouvoir fonctionner à bord d'un robot réel aux données sensorielles bruitées.

Ce choix s'est porté sur le système de navigation de Filliat (2001), qui ne prétend pas être un modèle biomimétique de navigation, bien que son codage par population de la localisation évoque l'activité des cellules de lieu de l'hippocampe. En revanche, ses fonctions sont semblables à celles que l'on attendrait d'un modèle plus biomimétique. Ses capacités de cartographie et de localisation émulent le fonctionnement de l'hippocampe, alors que sa capacité de planification est caractéristique du cortex préfrontal. Il a été testé en simulation, mais a également prouvé sa robustesse lors d'expériences embarquées à bord d'un robot Pioneer (©ActivMedia) dans l'environnement « naturel » du laboratoire.

4.5.2 Description

Ce système de navigation génère une carte topologique de l'environnement du robot et le localise sur cette carte. Pour cela, il utilise à la fois les données idiothétiques (mesure des déplacements propres) et allothétiques (vision panoramique, mesures sonar omnidirectionnelles).

Cette carte se présente sous la forme d'un graphe. Les nœuds du graphe représentent les lieux de l'environnement précédemment explorés. Ils stockent les informations allothétiques spécifiques à ces lieux. Ils sont reliés par des arêtes dont l'orientation et la longueur dépendent des données idiothétiques recueillies durant les déplacements entre nœuds.

La localisation du robot dans la carte passe par le calcul en chaque nœud de la probabilité que le robot soit présent en ce nœud. Cette probabilité est calculée en combinant, d'une part, une mesure de similitude entre ce qui est perçu et les données allothétiques de chaque nœud de la carte, d'autre part, une position déduite de la dernière localisation et des derniers déplacements effectués. La position du robot peut être déduite par le calcul du barycentre des nœuds pondérés par leur activité. Cependant, on considèrera également que le nœud le plus activé est le nœud courant.

La carte est construite en ligne : lorsqu'aucun nœud n'a une activité suffisante, le robot est supposé être sorti du domaine actuel de la carte et un nouveau nœud est alors créé. Afin que cette construction soit robuste, le robot alterne phases d'exploration et retour vers les zones bien connues. Enfin, l'utilisation d'une procédure de suppression des nœuds inutilisés, d'une procédure de fusion de nœuds proches et enfin d'un algorithme de relaxation du réseau permet d'assurer la cohérence et la simplicité de la carte.

On notera qu'en l'état actuel, ce système de navigation suppose la présence à bord du robot d'une mesure de la direction absolue (compas).

A partir d'une carte, même partielle, le système de navigation est capable de planifier des déplacements : si un nœud de la carte est désigné comme but à atteindre, un algorithme calcule en chaque nœud de la carte une politique (direction à suivre) permettant d'atteindre ce nœud. On

notera qu'une étape de la méthode de calcul de cette politique consiste, pour chaque nœud, en la discrétisation de l'ensemble des directions possibles suivant un pas de $10°$, suivie du calcul, pour chacune des 36 direction résultantes, d'une intensité. Plus cette intensité est élevée, plus la direction est recommandée pour atteindre le but. C'est donc la direction qui a l'intensité la plus forte qui est choisie comme politique en ce nœud.

Dans le cadre de l'intégration de ce système de navigation avec le modèle de sélection de l'action, c'est le résultat de cette étape intermédiaire (un *profi l de direction*) qui nous intéressera, plutôt que la politique résultante.

```
si perdu = vrai alors
    dir ← RetourSurSesPas();
    datePerdu ← date ;
sinon
    si butActif = vrai alors
        dir ← Planification(but);
    sinon
        si (date- datePerdu) > 4 alors
            dir ← Exploration();
        sinon
            dir ← RetourZoneCartographiée();
        fin
    fin
fin
```

Algorithme 1: Algorithme de sélection de la direction de déplacement du système de navigation de Filliat.

En l'absence de mécanisme de sélection de l'action, le contrôle du robot assure la construction d'une carte robuste en alternant de courtes phases d'exploration avec des retours réguliers dans les zones bien cartographiées. Ainsi, chaque fois que le robot a parcouru plus d'une certaine distance (50cm dans les expériences de Filliat) en ligne droite, il s'arrête, cherche à se localiser dans sa carte, la met à jour (création d'un nouveau nœud, fusion de plusieurs nœuds, etc.) et choisit une nouvelle direction suivant l'algorithme 1. On constate que ce choix de direction fait appel à quatre « comportements » de base :

- « retour sur ses pas » : propose un déplacement qui est l'inverse des dernières mesures odométriques,
- « retour en zone cartographiée » : la direction fournie dirige le robot vers la zone de la carte où la densité de nœuds est la plus forte,
- « exploration » : le robot est envoyé vers une zone libre d'obstacles et où la densité de nœuds est faible,

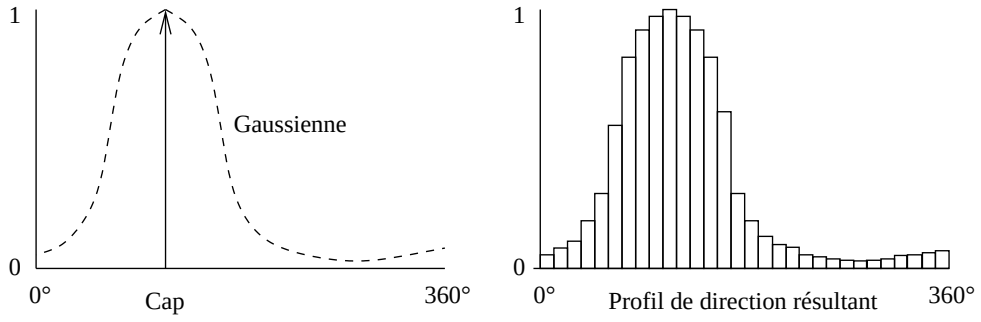

FIG. *4.10: Transformation d'un cap absolu en un* profil de direction.

– « planification » : si un nœud but est fourni, une direction permettant de se rapprocher de ce but est fournie.

Ce contrôle fait que la carte est construite petit à petit, ce qui en assure la cohérence. En revanche, il n'est absolument pas contraint par des impératifs métaboliques. L'intégration de la navigation et des contraintes de survie envisagée ici va donc nécessiter de prendre en compte les éventuels conflits entre la construction chronique de la carte décrite ci-dessus et la recherche, parfois dans l'urgence, des ressources consommables de l'environnement.

4.5.3 Adaptation du système de navigation pour l'interface

Le modèle tel qu'il a été décrit ci-dessus a subi quelques modifications afin d'être compatible avec notre interfaçage. Le système de navigation n'ayant plus le contrôle direct du robot, il ne peut que recommander les directions qui lui conviennent au système de sélection de l'action, qui décide d'en tenir compte ou non. Les quatre « comportements » évoqués ci-dessus (retour sur ses pas, retour en zone cartographiée, exploration et planification) seront donc transformés en *profils de direction* fournis en entrée du système de sélection de l'action.

Des caps aux profils de direction

Si on laisse de côté pour l'instant la planification, détaillée plus bas, le système de navigation est susceptible de suggérer au système de sélection de l'action trois directions de mouvement correspondant aux trois comportements précédemment évoqués : « retour sur ses pas », « exploration » et « retour vers une zone cartographiée ». Ces suggestions prennent chacune la forme d'un cap à suivre. Afin de les rendre compatible avec la sélection effectuée par la boucle ventrale, on les transforme en trois *profils de direction*, respectivement **RSP**, **Exp** et **RZC**.

Les composantes des *profils de direction* sont calculées à partir d'une gaussienne centrée sur le cap. Ainsi, pour un comportement donné, l'intensité associée à une direction est d'autant plus forte que cette direction est proche du cap suggéré par le système de navigation (fig. 4.10).

Pour transformer le cap fourni par un comportement cpt donné, cap_{cpt}, en une intensité pour une direction donnée, $I_{cpt}(dir)$, on calcule tout d'abord un écart au cap, $\Delta(dir)$ auquel on applique ensuite la gaussienne :

$$\Delta(dir) = \begin{cases} \mid cap_{cpt} - dir \mid /180 & \text{si } \mid cap_{cpt} - dir \mid < 180, \\ (360 - \mid cap_{cpt} - dir \mid)/180 & \text{sinon.} \end{cases} \tag{4.5}$$

$$I_{cpt}(dir) = \exp(-20 \times \Delta(dir)^2) \tag{4.6}$$

Désorientation

La notion de désorientation a du également être assouplie afin que le système de sélection de l'action puisse de lui-même choisir quand suivre les recommandations issues de l'un des trois *profi ls de direction* suggérés par le système de navigation. Le système de navigation produit donc à chacune de ses mises à jour une variable Des, comprise entre 0 et 1, mesurant le degré de désorientation du robot. Elle sera considérée par le système de sélection de l'action comme une variable interne du robot. Elle est construite de sorte à augmenter fortement lorsque le robot sort d'une zone connue, pour atteindre son maximum au bout de 16 mises à jour du système de navigation. Inversement, elle décroît lorsque l'on revient dans une zone connue, pour revenir à 0 au bout de 4 mises à jour du système de navigation.

Elle est calculée comme suit :

– Si l'on est dans une zone inconnue depuis n pas de temps, elle vaut :

$$Des = \begin{cases} 0,5 + 0,5 \times n/16 & \text{si } n < 16, \\ 1 & \text{sinon.} \end{cases} \tag{4.7}$$

– sinon, si l'on est de retour dans une zone connue depuis n pas de temps, elle vaut :

$$Des = \begin{cases} 0,5 \times (4 - n)/4 & \text{si } n < 4, \\ 0 & \text{sinon.} \end{cases} \tag{4.8}$$

Afin de simplifier les choix opérés par la sélection de l'action et rester proche de l'algorithme original de contrôle du robot, la boucle NAcc « core » ne reçoit en réalité que le *profi l de direction* de l'« exploration », $\mathbf{Exp}$, et un second *profi l de direction* permettant de « retrouver son chemin » ($\mathbf{RSC}$), utilisant l'odométrie (profil de « retour sur ses pas » $\mathbf{RSP}$) si la désorientation est forte et la carte (profil de « retour en zone connue » $\mathbf{RZC}$) sinon :

– si $Des > 0,5$:

$$\mathbf{RSC} = \mathbf{RSP} \tag{4.9}$$

– sinon :

$$\mathbf{RSC} = \mathbf{RZC} \tag{4.10}$$

Planification et ressources

Les capacités de localisation et de planification du système de navigation peuvent être exploitées pour permettre au robot de retrouver des ressources de l'environnement situées hors de son champ perceptif. A cet effet, nous avons ajouté la possibilité d'associer par apprentissage certains nœuds de la carte aux types de ressources susceptibles de s'y trouver. L'association entre un nœud i de la carte et un type de ressources res est caractérisée par une force $W(res, i)$, initialisée à 0 et comprise entre 0 et 1. Elle est d'autant plus grande que la ressource a effectivement été trouvée lors des passages du robot à cet endroit. Inversement, s'il repasse par un nœud associé à une ressource et qu'elle ne s'y trouve pas, la force de l'association décroît.

Afin de permettre un apprentissage rapide et localisé, lorsqu'une ressource de type res est détectée, seul le nœud courant nc (c'est-à-dire celui qui est la localisation la plus probable du robot) voit la force de son association avec ce type de ressources, $W(res, nc)$, augmenter :

– si res est observée en nc pour la première fois :

$$W(res, nc)(t) = 0,8 \tag{4.11}$$

– si res a déjà été observé en nc :

$$W(res, nc)(t) = W(res, nc)(t-1) + 0,8 \times (1 - W(res, nc)(t-1)) \tag{4.12}$$

Au contraire, lorsqu'une ressource a disparu, on veut pouvoir oublier sa présence sur tous les nœuds qui lui sont proches, sans que le robot soit obligé de les visiter un à un. L'oubli est donc appliqué à tous les nœuds i qui ont une probabilité $prob(i)$ de présence du robot non-nulle :

– si $W(res, i)(t-1) \neq 0$ et que res n'est pas observée :

$$W(res, i)(t) = W(res, i)(t-1) - 0,2 \times prob(i) \times W(res, i)(t-1) \tag{4.13}$$

L'apprentissage est donc localisé et rapide, alors que l'oubli est plus lent, mais réparti.

Les nœuds associés à la présence de ressources étant caractérisés par ces poids, il est possible d'utiliser l'algorithme de planification de chemin pour générer des suggestions de déplacement permettant de retrouver un type de ressources donné. Ainsi que cela a été précisé plus haut, on s'intéresse non à l'étape finale des calculs de planification –une unique direction en chaque nœud– mais au résultat intermédiaire : des intensités associées à 36 directions espacées d'un pas constant de 10°, c'est-à-dire un *profi l de direction*.

La méthode utilisée à cet effet est la suivante : pour chaque ressource res, on applique la planification à tous les nœuds i de la carte dont le poids $W(res, i)$ n'est pas nul. On obtient alors pour chacun de ces nœuds un vecteur de 36 intensités $\mathbf{P}(\mathbf{res}, \mathbf{i})$. On pondère ce vecteur par le poids $W(res, i)$ de sorte qu'un nœud faiblement associé à la ressource ne génère que de faibles intensités.

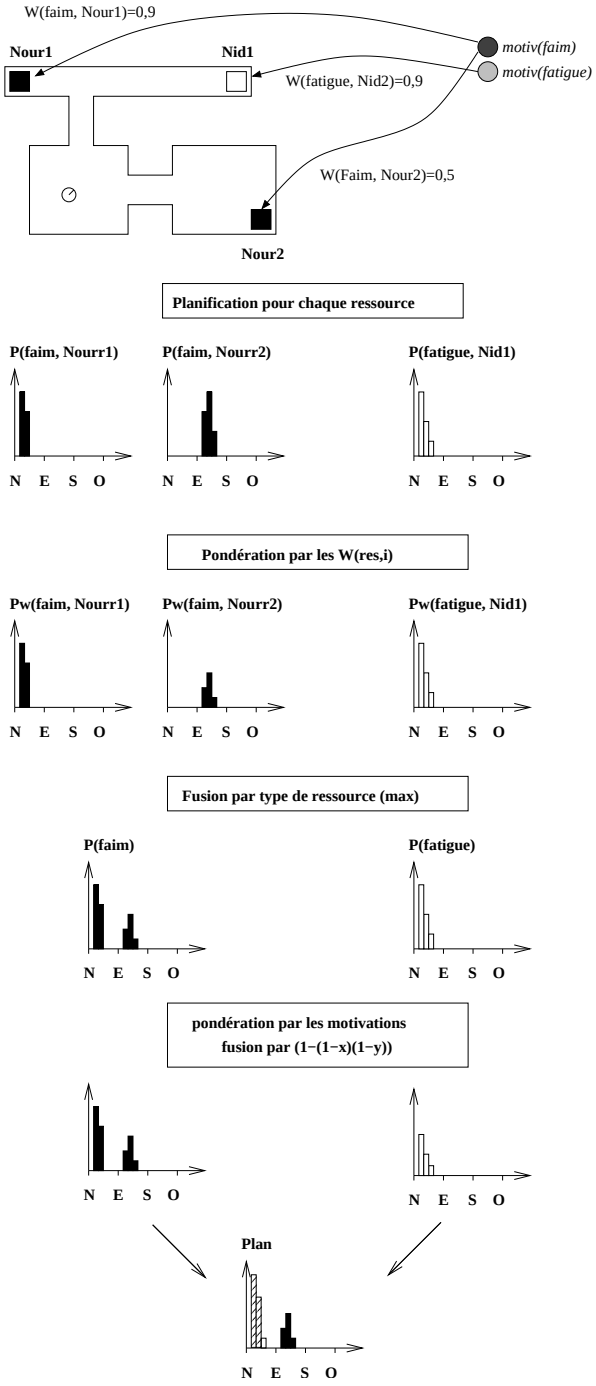

FIG. *4.11: Récaputulation des calculs de planification qui, partant d'une situation environnementale et motivationnelle donnée aboutissent au* profil de direction **Plan**. *Détail des notations dans le texte.*

$$\mathbf{Pw}(\mathbf{res}, \mathbf{i}) = W(res, i) \cdot \mathbf{P}(\mathbf{res}, \mathbf{i}) \tag{4.14}$$

Les vecteurs pondérés $\mathbf{Pw}(\mathbf{res}, \mathbf{i})$ résultants sont alors combinés par l'application de l'opérateur max, composantes par compostantes. On obtient au final autant de vecteurs de planification $\mathbf{P}(\mathbf{res})$ qu'il y a de type de ressources dans l'environnement, vecteurs qui sont fournis au système de sélection de l'action.

$$\mathbf{P}(\mathbf{res}) = \begin{pmatrix} \max_i(\mathbf{Pw}(\mathbf{res}, \mathbf{i})_1) \\ \vdots \\ \max_i(\mathbf{Pw}(\mathbf{res}, \mathbf{i})_{36}) \end{pmatrix} \tag{4.15}$$

Enfin, ces vecteurs $\mathbf{P}(\mathbf{res})$ vont être combinés en un seul *profi l de direction* final, $\mathbf{Plan}$, représentant les suggestions de déplacement issues de la planification sur chaque type de ressource. Pour ce faire, chaque $\mathbf{P}(\mathbf{res})$ va être pondéré par la « motivation » à rechercher les ressources de type res. Nous parlerons ici de « motivation » plutôt que de variables internes. En effet, nous avons pu constater dans les expériences du chapitre précédent que l'intérêt porté à certaines ressources peut dépendre d'une combinaison de variables internes, que nous appellerons donc « motivation ». En l'espèce, la salience de $RechargeSurBlanc$ dépendait non du manque d'*Energie*, mais d'une combinaison de ce manque et des réserves d'*Energie Potentielle* ($Circ(Rev(E_{Pot})) \times Rev(E)$, voir tab. 3.3). $\mathbf{Plan}$ est donc calculé comme une combinaison des $\mathbf{P}(\mathbf{res})$ pondérés par leur « motivation » correspondante. Cette combinaison sera effectuée via un opérateur de type $1 - [(1 - x)(1 - y)]$ qui permet, contrairement au max, de donner un avantage à une direction favorisée par plusieurs *profi ls de direction* vis-à-vis d'une direction favorisée par un seul d'entre eux.

$$\mathbf{Plan}_i = 1 - \prod_{res}(1 - motiv \times \mathbf{P}(\mathbf{res})_i) \tag{4.16}$$

L'ensemble de ces calculs de planification sont récapitulés en figure 4.11, dans un exemple illustratif où les motivations concernent la fatigue et la faim.

Le système de navigation, suite à ces modifications, fournit donc à la boucle NAcc « core » (fig. 4.12) :

- trois profils de direction correspondant aux suggestions des comportements « retrouver son chemin » (**RSC**) et « exploration » (**Exp**), et aux suggestions de déplacement permettant d'atteindre par planification les ressources nécessaires à un instant donné (**Plan**) (fig. 4.12).
- une variable interne représentant le degré de désorientation (Des),

4.6 Expérimentations dans une tâche de survie en environnements simples

L'évaluation de l'adaptativité des comportements générés par ce modèle intégrant navigation et sélection de l'action sera effectuée uniquement en simulation, pour un animat disposant d'informations plus riches que lors des expériences de sélection de l'action seule, puisqu'il est en particulier capable de détecter les ressources de l'environnement à distance. Cette évaluation sera fondée sur la résolution d'une tâche de survie similaire à celle utilisée au chapitre précédent.

Cette évaluation est réalisée en deux temps.

Tout d'abord, l'animat sera testé dans des labyrinthes simples, dans des conditions contrôlées, afin de tester son comportement dans cinq situations types. On évaluera tout d'abord l'efficacité de la stratégie de navigation topologique, puis on s'intéressera aux effets induits par l'apparition ou la disparition de ressources. Puis deux expériences permettront d'évaluer la capacité du système à effectuer des choix rationnels en fonction de la configuration environnementale et de l'état interne.

Enfin, sa capacité à survivre en exploitant les propriétés mises en évidence dans ces cinq premières expériences sera testée dans un environnement complexe, sur une longue durée.

4.6.1 Matériel et méthode

Environnement

Les environnements (en deux dimensions) sont constitués de segments pouvant prendre 256 teintes de gris, depuis le noir jusqu'au blanc. Ils constituent des obstacles et leur teinte sert de repère visuel au système de navigation.

Les ressources affectant la survie de l'animat dans ces environnements sont ici de trois types différents. Elles sont représentées par des obstacles carrés de 50cm de côté. Leur teinte spécifique permet d'en identifier le type : gris médian (teinte 127) pour les sources d'« ingestion », blanc (teinte 255) pour les zones de « digestion » et enfin gris sombre (teinte 31) pour les *Zones Dangereuses*. Le robot est considéré à portée d'une de ces ressources si elle occupe plus de 60° de son champ visuel, ce qui implique une distance maximale entre le centre du robot et le centre de la ressource d'environ 70cm. Le processus de vision est simplifié à l'extrême, puisqu'aucune notion d'éclairage n'est prise en compte : les senseurs visuels ont directement accès à la teinte d'un objet.

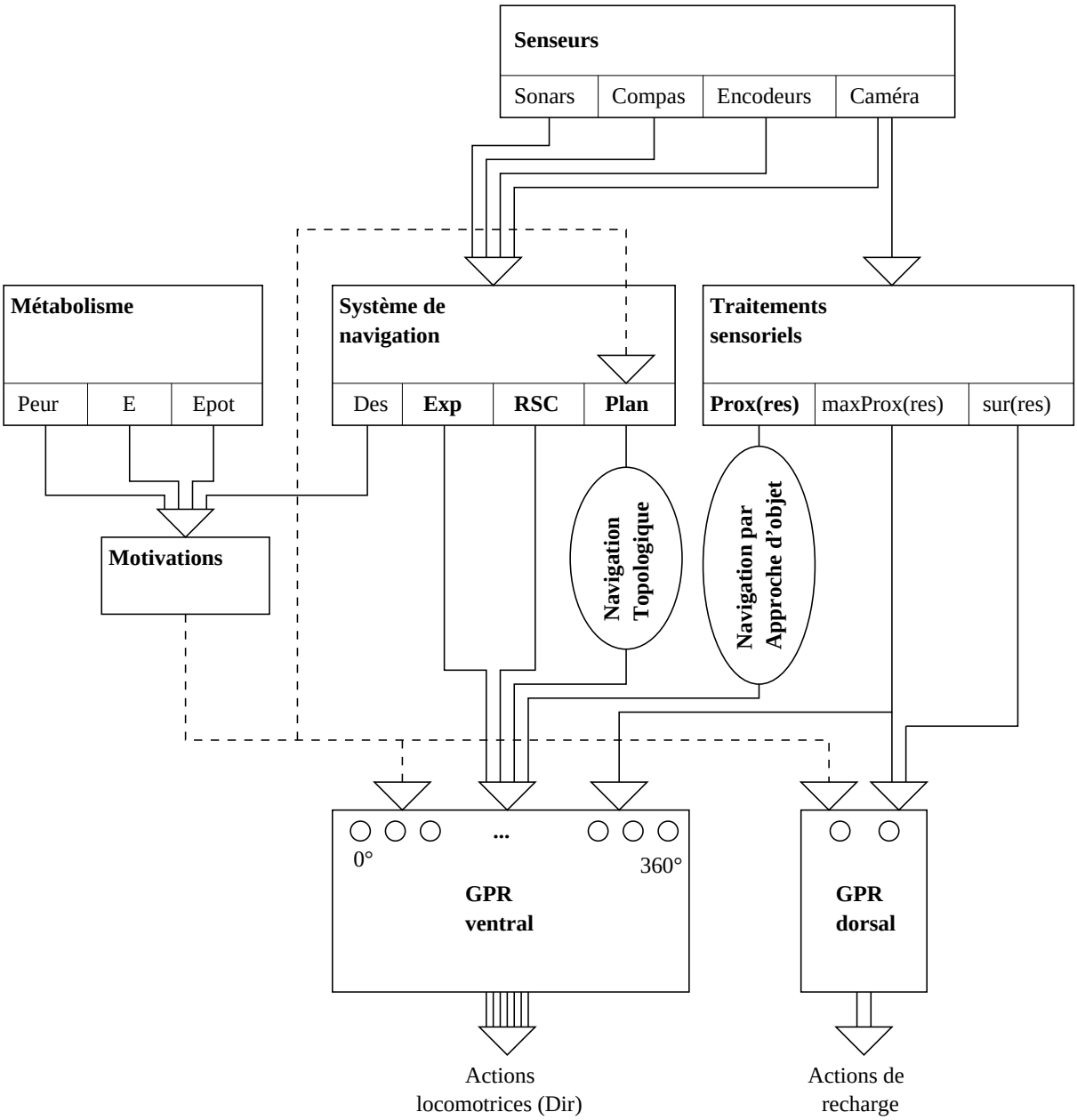

FIG. *4.12: Organisation des entrées internes et externes des deux boucles de sélection. Les profils de direction sont en gras, les motivations sont représentées par une flèche pointillée. Détails dans le texte.*

Robot simulé

Le robot simulé est de forme circulaire (30cm de diamètre) et est équipé de deux roues motrices lui permettant de tourner sur lui-même, sa vitesse en translation est de $40cm.s^{-1}$ et, en rotation, de $10°.s^{-1}$. Ses senseurs simulés sont :

- une caméra linéaire panoramique de faible résolution : un pixel par secteur de 10°,
- une ceinture de huits sonars : portée maximale de $5m$, incertitude sur la direction pointée de plus ou moins $5°$, erreur de mesure entre 0 et $20cm$),
- des encodeurs permettant de mesurer les déplacements propres : erreur de plus ou moins 5% vis-à-vis de la distance réellement parcourue,
- un compas : erreur de plus ou moins $10°$ vis-à-vis de la direction réelle.

Variables externes

Le système de sélection de l'action ne tient compte, du point de vue des senseurs, que des données fournies par la caméra. L'image de la caméra (une ligne de 36 pixels) est soumise à divers traitements afin de fournir au système de sélection de l'action neuf variables sensorielles (trois par type de ressources).

Les trois premières sont les *profi ls de direction* correspondant à la navigation par approche d'objets : $\mathbf{Prox(31)}$, $\mathbf{Prox(127)}$ et $\mathbf{Prox(255)}$. Elles viennent compléter les *profi ls de direction* fournis par le système de navigation ($\mathbf{P}$ et $\mathbf{RSC}$), à partir desquels la boucle ventrale va effectuer ses choix de direction de locomotion (fig. 4.12).

Les six autres sont : $maxProx(31)$, $maxProx(127)$, $maxProx(255)$, $sur(31)$, $sur(127)$ et $sur(255)$. Elles seront principalement utilisée par la boucle dorsale (fig. 4.12).

Les mesures obtenues par les sonars, les encodeurs et le compas ne sont utilisées que par le système de navigation (fig. 4.12).

Les *profi ls de direction* $\mathbf{Prox(31)}$, $\mathbf{Prox(127)}$ et $\mathbf{Prox(255)}$ permettent l'approche d'objets. Ce sont en effet des vecteurs de 36 mesures, donnant une approximation grossière de la proximité de chaque type de ressources dans chaque direction. Ces mesures sont proportionnelles au nombre de pixels de la teinte de la ressource adjacents autour de chaque direction. Considérant une teinte de ressource donnée res, la composante du profil $\mathbf{Prox(res)}$ dans la direction i, est proportionnelle au nombre de pixels de teinte res dans une fenêtre de 7 pixels centrée sur i. La valeur de maximale, 1, est atteinte pour un bloc d'au moins 6 pixels :

$$\mathbf{Prox(res)}_i = \begin{cases} 1 & \text{si } temp > 1, \\ temp & \text{sinon.} \end{cases} \tag{4.17}$$

avec $temp$ défini par :

$$temp = \frac{1}{6}\sum_{j=i-3}^{j=i+3} \delta_{res,Img(j)} \tag{4.18}$$

et $\delta_{i,j}$ (symbole de Krönecker) :

$$\delta_{res,Img(i)} = \begin{cases} 1 & \text{si } res = Img(i), \\ 0 & \text{sinon.} \end{cases} \tag{4.19}$$

Les trois variables entières suivantes sont les maximums de ces profils de direction. Il s'agit de $maxProx(31)$, $maxProx(127)$ et $maxProx(255)$, trois variables non-directionnelles dénotant la proximité de chaque type de sources :

$$maxProx(res) = \max_{i\in[1,36]} \mathbf{Prox}(\mathbf{res})_i \tag{4.20}$$

Enfin, les trois dernières sont des booléens $sur(31)$, $sur(127)$ et $sur(255)$ qui permettent à l'animat de savoir quand il est assez près d'une source pour y avoir accès :

$$sur(res) = \delta_{1,maxProx(res)} \tag{4.21}$$

Variables internes

Le robot est doté de quatre variables internes : E, E_{Pot}, Des et $Peur$ (fig. 4.12).

Le métabolisme virtuel implémenté dans cette expérience est du même type que celui utilisé dans les expériences de survie (voir 3.2.1) : le robot possède deux variables internes liées à sa survie E et E_{Pot}, à valeurs dans [0,1]. Il « meurt » lorsque E descend à 0. Le robot consomme en permanence de l'*Energie*, au taux de $5.10^{-4} unite.s^{-1}$. Il doit donc s'en procurer par le mécanisme en deux étapes suivant :

1. « Ingestion » d'E_{Pot} : cela nécessite d'être sur une source d'E_{Pot}, de stopper et de sélectionner en boucle dorsale le comportement $RechargeE_{Pot}$. Dans ces conditions, la consommation d'*Energie* reste de $5.10^{-4} unite.s^{-1}$, en revanche le gain d'*Energie Potentielle* est de $3.10^{-2} unite.s^{-1}$:

$$\Delta E_{Pot} = 3.10^{-2} \times T_{Ingest} \tag{4.22}$$

2. « Digestion » de l'E_{Pot} en E : cela nécessite d'être sur une source d'E, de stopper et de sélectionner en boucle dorsale le comportement $RechargeE$. Tant que le robot a une réserve suffisante d'E_{Pot}, elle est transformée en *Energie* au taux de 3.10^{-2} unité/sec :

$$\Delta E_{Pot} = -3.10^{-2} \times T_{Digest} \tag{4.23}$$

$$\Delta E = 3.10^{-2} \times T_{Digest} \tag{4.24}$$

Un animat initialisé avec une *Energie* à 1 et n'exécutant aucune recharge a une durée de vie de 33 min ($2000s$).

Les *Zones Dangereuses* présentes dans l'environnement sont susceptibles d'affecter l'*Energie* de l'animat : si une zone dangereuse est proche ($maxProx(31) > 0.5$), il a $(maxProx(31) \times 10)\%$ de chances de perdre 0.1 d'*Energie* à chaque mise à jour du système de sélection de l'action.

L'influence que les *Zones Dangereuses* exercent sur le robot, et en particulier sur les processus de planification, est modulée par une variable de *Peur*, dont la valeur est constante pour la durée d'une expérience.

Une dernière variable interne mesurant le degré de désorientation de l'animat, Des, est fournie par le système de navigation (cf. 4.5.3). Cette variable n'étant mise à jour qu'épisodiquement (cf. *Ordonnancement des calculs*, plus bas), c'en est une version lissée, Des_L, qui est utilisée au niveau de la sélection de l'action. Elle est initialisée à 0 en début d'expérience et tend progressivement vers Des à chaque cycle du système de sélection de l'action :

$$Des_L(t) = 0,9 \times Des_L(t-1) + 0,1 \times Des \tag{4.25}$$

Motivations

Au vu des expériences réalisées au chapitre précédent, il apparaît que les comportements liés à l'approche et la consommation d'une ressource doivent être pondérés par une combinaison de variables internes spécifiques, que nous nommerons « motivation ».

Dans ces précédentes expériences, il est apparu que la motivation associée à la recharge en *Energie Potentielle* était calculée par $Rev(E_{Pot})$ et que celle associée à la transformation de l'*Energie Potentielle* en *Energie* l'était par $(Circ(Rev(E_{Pot})) \times Rev(E))$.

Nous utiliserons donc ces deux motivations intermédiaires pour moduler les comportements de recharge. Les comportements permettant de retrouver son chemin et d'éviter les *Zones Dangereuses* le seront par les valeurs brutes de, respectivement, Des_L et $Peur$, que nous considèrerons également comme des motivations par commodité de notation :

$$\begin{aligned} motiv(E) &= (Circ(Rev(E_{Pot})) \times Rev(E)) \\ motiv(E_{Pot}) &= Rev(E_{Pot}) \\ motiv(RSC) &= Des_L \\ motiv(ZD) &= Peur \end{aligned} \tag{4.26}$$

Ces quatre « motivations » sont utilisées en boucle ventrale pour moduler la planification (voir equ. 4.16) et l'approche d'objets (voir les calculs de saliences en annexe C.3), ainsi qu'en boucle dorsale pour moduler la salience des comportements de recharge (voir les calculs de salience en annexe C.3).

Comportements

La boucle dorsale sélectionne deux comportements : $RechargeE_{Pot}$ et $RechargeE$.

La boucle ventrale sélectionne des directions de locomotion. Cette sélection s'appuie sur six *profi ls de direction* :

1. $\mathbf{RSC}$, les suggestions de déplacement permettant de « retrouver son chemin » issues du système de navigation,
2. $\mathbf{Exp}$, les suggestions de déplacement d'« exploration » issues du système de navigation,
3. $\mathbf{Plan}$, les résultats combinés de l'ensemble des opérations de planification, pondérées par les « motivations »,
4. à 6. $\mathbf{Prox}(\mathbf{res})$, les trois profils de direction fournis par le traitement de l'image de la caméra permettent d'exhiber des comportements d'approche (*Energie* ou *Energie Potentielle*) ou d'évitement (*Zones Dangereuses*) des ressources visibles.

Enfin, l'évitement d'obstacles est implémenté sous forme d'un « réflexe ». Il n'est pas pris en charge par le système de sélection de l'action, mais directement par le programme bas niveau interfaçant la sélection de l'action et le contrôle du robot : si la direction demandée par la sélection de l'action amène le robot trop près d'un obstacle, elle est ignorée au profit d'une direction de contournement. L'utilisation de circuits courts « réflexe » permettant des réactions beaucoup plus rapides que celle passant par les circuits longs des ganglions de la base est en accord avec les données biologiques concernant l'amygdale (Rolls, 1999). Cependant, ces circuits « réflexes » ne sont pas ici issus d'une modélisation biomimétique.

Ordonnancement des calculs

Les systèmes de navigation et de sélection de l'action ne sont pas mis à jour à la même fréquence. En effet, pour ne pas sur-échantillonner la carte, le système de navigation ne doit être mis à jour que lorsque que le robot s'est éloigné de plus de $50cm$ de sa dernière position.

Une mise à jour du système de navigation correspond au repositionnement de l'animat dans la carte, à l'ajout, la suppression ou la fusion de nœuds si cela s'avère nécessaire, à la mise à jour des forces associant motivations et nœuds et enfin au recalcul des profils de direction correspondant à la planification, l'exploration et le retour en zone connue. Ces calculs de navigation modifient de nombreux *profi ls de direction* utilisés par le système de sélection de l'action. Il est donc autorisé lors de chaque mise à jour du système de navigation, à effectuer 30 mises à jour, afin de stabiliser sa sélection. On appellera cette séquence d'une mise à jour de la navigation et de 30 mises à jour de la sélection de l'action une *mise à jour longue* du modèle d'intégration, elle dure en général de l'ordre de $2s$.

Entre deux mises à jour de la navigation, tant que $50cm$ n'ont pas été parcourus, l'animat continue de recevoir de nouvelles données de la caméra et de modifier les variables qui en découlent, alors que les *profi ls de direction* issus du système de navigation, eux, ne changent pas. A chacune de ces réceptions, l'information changeant peu, 6 mises à jour du système de sélection de l'action sont autorisées. Il s'agira d'une *mise à jour courte* (environ $0,8s$) du modèle d'intégration.

Dans un cas comme dans l'autre, des ordres sont envoyés au simulateur. Ils résultent des inhibitions de sortie du modèle des ganglions de la base à la dernière des mises à jour autorisées. On notera qu'en moyenne, en tenant compte des deux types de mises à jour, une mise à jour du modèle d'intégration a lieu toutes les $0,87s$.

4.6.2 Expérimentations et résultats

Expériences

Les cinq premières expériences testent le comportement du modèle dans des situations type. Elles se déroulent dans des environnements simplifiés, limités aux seules caractéristiques nécessaires pour l'évaluation. Les niveaux initiaux des variables internes sont fixés selon les besoins des expériences. Les calculs de saliences sont susceptibles de varier d'une expérience à l'autre, d'une part pour souligner les effet induits par la variation d'un paramètre, d'autre part pour l'améliorer en fonction des limitations rencontrées. Le détail des calculs de saliences pour chaque expérience figure en annexe C.3.

A l'exception de la première, toutes ces expériences utilisent des cartographies de l'environnement préalablement établies par le robot lors d'une phase d'apprentissage. Durant ces phases d'apprentissage, le robot est libéré de toute contrainte métabolique et est autorisé à cartographier l'environnement durant deux à trois heures (selon la taille de l'environnement) le temps qu'il connaisse tout l'environnement, c'est-à-dire que sa variable de désorientation soit nulle où qu'il se trouve.

Les données concernant les variables externes, internes, les saliences, les valeurs des inhibitions en sortie des deux boucles, la direction de déplacement et l'éventuel comportement de recharge sélectionnés sont enregistrées à chaque mise à jour du modèle (longue ou courte).

L'expérience 1 évalue l'efficacité d'une stratégie de navigation, la planification topologique. Le robot est donc placé dans un environnement comprenant une ressource d'*Energie* et une ressource d'*Energie Potentielle* non-visibles simultanément et éloignées l'une de l'autre.

L'expérience 2 s'intéresse à la capacité du système à coordonner deux stratégies de navigation,

la planification et l'approche d'objets, en faisant preuve d'opportunisme. Il aura à se détourner d'un chemin planifié pour rejoindre une ressource visible nouvellement apparue dans l'environnement.

L'expérience 3 évalue de la même façon la capacité du robot à s'adapter aux changements de l'environnement lors de la disparition d'une ressource.

Les expériences 4 et 5 comparent le comportement du modèle selon l'état interne de l'animat lorsqu'une zone dangereuse placée sur un chemin court l'oblige à choisir entre ce chemin et un détour (exp. 4) ou lorsqu'il doit choisir entre un chemin court n'aboutissant qu'à un type de ressources et un chemin long aboutissant aux deux types de ressources (exp. 5).

Les expériences précédentes sont destinées à évaluer de façon quantitative chacune des capacités adaptative que l'interfaçage navigation/sélection de l'action peut apporter à un animat. Certaines, cependant, n'ont pas été spécifiquement conduites dans le cadre d'une expérience de survie (exp. 2 à 5). C'est pourquoi une dernière expérience consiste à placer l'animat dans un environnement de grande taille où tous les problèmes posés isolément précédemment sont réunis, où tous les types de ressources sont présents et dans laquelle sa tâche sera de survivre le plus longtemps possible. A l'instar de l'expérience 1, un animat sans système de planification sera également placé dans les mêmes conditions.

Expérience 1 : utilisation de la cartographie et de la planification

Les tests de l'expérience 1 ont été menés dans l'environnement représenté par la figure 4.13. Il mesure $9m \times 7m$ et comporte deux ressources, l'une d'*Energie* et l'autre d'*Energie Potentielle*. Le robot est initialement placé à proximité de la ressource d'*Energie*. Les deux ressources sont hors de vue l'une de l'autre, éloignées et séparées par une série d'embranchements et de couloirs, de sorte qu'un robot réactif, capable d'approche d'objets visibles mais pas de planification, ait des difficultés à y survivre longuement.

Résultats

Dix tests avec un système de navigation actif et dix tests sans ont été effectués. Nous avons choisi comme critère de réussite du test la capacité à survivre plus de 4 heures. Les médianes des durées des deux ensembles de tests sont comparées par le test U de Mann-Whitney.

Sur dix essais, le robot réactif s'est avéré incapable de survivre plus de $2h30$ et donc de résoudre la tâche (voir fig. 4.14). Le robot doté d'un système de cartographie et de planification a survécu significativement plus longtemps que le réactif (voir tab. 4.1), il n'a cependant été capable de

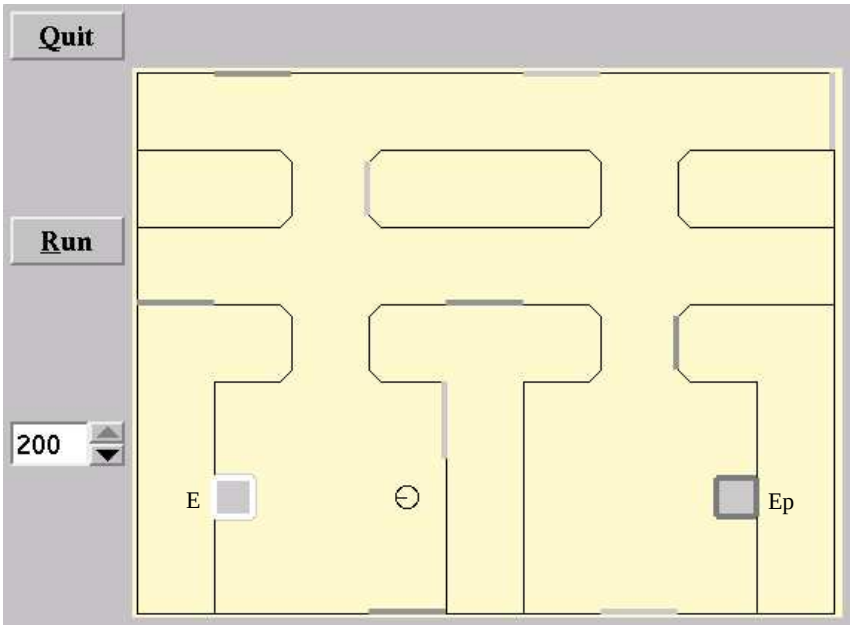

FIG. *4.13: Expérience 1 : environnement de test comportant une ressource d'*Energie *et une ressources d'*Energie Potentielle *éloignées et hors de vue l'une de l'autre. Le robot est positionné à l'emplacement initial commun à tous les essais.*

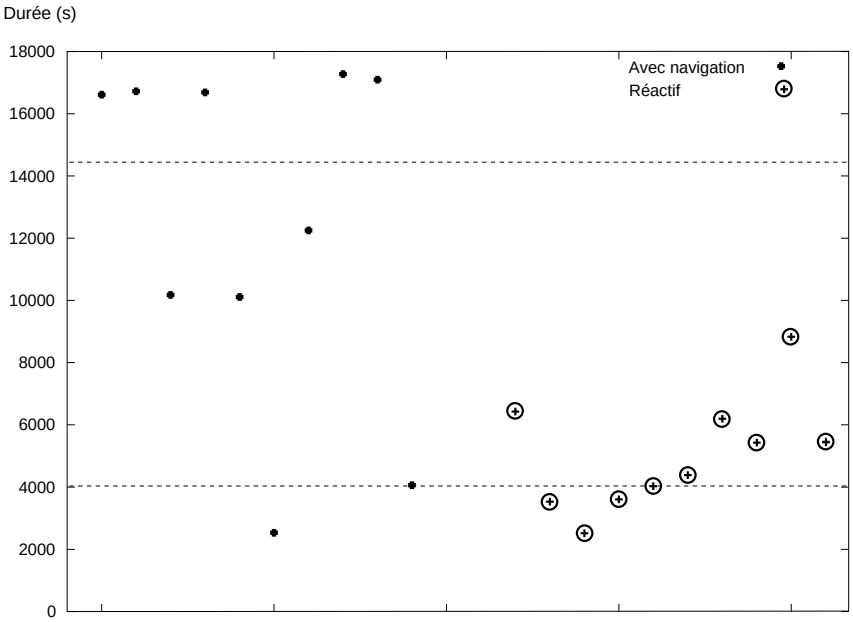

FIG. *4.14: Expérience 1 : durées (en secondes) des essais avec et sans système de navigation. Ligne pointillée inférieure : durée correspondant à la simple utilisation des réserves initiales ($4000s$). Ligne pointillée supérieure : limite des 4 heures de survie.*

Durées	Médiane	Intervalle
Navigation	14431,5	2531 :17274
Sans Navigation	4908	2518 :8831
U test	**U**	**p**
	15	p<0.01

TAB. *4.1: Expérience 1 : comparaison (test U de Mann-Whitney) entre les durées de vie du système utilisant le systèmle de navigation pour planifier ses déplacements et le système purement réactif.*

résoudre la tâche (survivre plus de $4h$) que cinq fois sur dix.

Les deux essais ayant duré moins de $4100s$ correspondent à des cas où le hasard induit par l'exploration du robot a fait qu'il n'arrive pas à trouver et enregistrer dans sa carte la localisation de la ressource d'*Energie Potentielle*. Dans ce cas[2] il ne peut que consommer sa charge initiale d'*Energie* puis transformer sa réserve initiale d'*Energie Potentielle* en *Energie* avant de mourir. La survenue de ce type d'échec est inévitable et ne dénote en rien un mauvais fonctionnement du système. Elle serait d'ailleurs plus fréquente si la distance et les ramifications entre les deux ressources étaient plus importantes.

En ce qui concerne les trois essais ayant duré entre $10.000s$ et $12.500s$, ils s'avère que le robot est mort alors qu'il était à proximité de la source d'*Energie Potentielle* et en grand manque d'*Energie Potentielle*, mais qu'une trop forte excitation en provenance de la sortie de la boucle ventrale par la boucle dorsale via la voie trans-subthalamique a provoqué un arrêt du robot. L'excitation en provenance de la boucle dorsale est due à une « anticipation » de l'*Energie Potentielle* codée par la dépendance de la salience de $RechargeE_{Pot}$ à la variable $maxProx(127)$ (voir tab. C.4, en annexe), qui permet de sélectionner plus rapidement ce comportement lorsque le robot atteint la ressource (de manière tout à fait similaire à l'« anticipation » due à la persistance, évoquée en 3.3). Le poids associé à $maxProx(127)$ est trop important, puisque lorsque E_{Pot} atteint des valeurs de l'ordre de $0,1$, ce blocage face à la ressource survient. Il devra donc être diminué pour les expériences suivantes.

Discussion

Cette première expérience montre que le système de navigation (cartographie en-ligne et planification) fonctionne correctement, puisqu'il donne au robot les avantages attendus dans un environnement où être réactif ne suffit pas. Le robot s'est avéré capable d'explorer et son environnement afin de trouver les deux types de ressources nécessaires à sa survie, puis d'exploiter

[2]La durée de vie théorique dans ces situation est de $4000s$ mais des erreurs d'arrondi pouvant survenir lors des calculs de mise à jour du métabolisme, on constate que le robot peut parvenir à survivre jusqu'à $4086s$

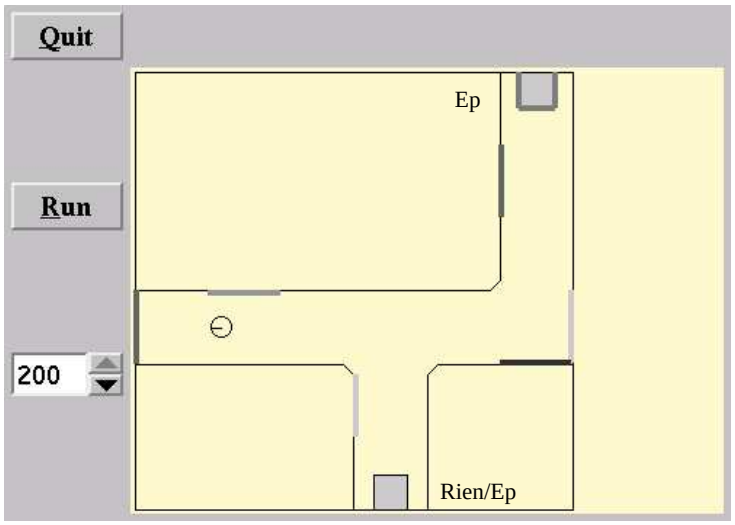

FIG. *4.15: Expérience 2 : environnement de test comportant une ressource permanente d'*Energie Potentielle *fixe (en haut), et une autre absente lors de la phase d'apprentissage et présente lors des essais (en bas). Le robot est positionné à l'emplacement initial commun à tous les essais.*

la cartographie résultante pour atteindre rapidement ces ressources par planification. Il apparaît cependant que les calculs de saliences choisis pour cette expérience doivent être modifiés.

Le blocage dû à la forte excitation de la voie trans-subthalamique semble pouvoir être résolu par un ajustage à la main des paramètres du calcul des saliences, mais ne peut être naturellement réalisé qu'après le constat de ces blocages, ici intervenu après près de $10h$ d'expérimentations. L'intégration de capacités d'adaptation en ligne comme, par exemple, de l'apprentissage par renforcement pourrait permettre au système de s'ajuster automatiquement dans ce genre de situation et le rendre conséquemment plus autonome. Ce point sera repris en discussion générale.

Expérience 2 : apparition de ressources et opportunisme

Les essais de l'expérience 2 ont été menés dans l'environnement représenté figure 4.15. Il mesure $6 \times 6m$ et comporte deux ressources d'*Energie Potentielle*, l'une est permanente, alors que l'autre est absente lors de la phase d'apprentissage de la carte mais peut être présente lors des essais. Le robot est positionné initialement à l'extrémité gauche du labyrinthe (voir fig. 4.15), en situation de manque d'*Energie Potentielle* ($E = 1$ et $E_{Pot} = 0,5$). Les essais sont arrêtés dès que le robot active $RechargeE_{Pot}$ sur l'une ou l'autre des ressources.

Résultats

Trois séries de quinze tests, correspondant à divers paramétrages du calcul de saliences, sont comparées, via le décompte du nombre d'occurrences du détour opportuniste. La comparaison

Pondération		Choix	
Planif.	**Vision**	**Haut**	**Bas**
0,65	0,55	13	2
0,55	0,55	7	8
0,45	0,55	2	13

TAB. *4.2: Expérience 2 : décompte des choix opérés selon la pondération associée à chacune des stratégies de navigation.*

de ces décomptes étant appliquée à un petit effectif c'est le test exact de Fisher qui est utilisé.

Lors des essais témoins, seule la source fixe dont la localisation a été apprise durant la phase d'apprentissage est présente. Le comportement du robot est semblable dans les dix essais : il se dirige directement vers cette source, sans faire de détours, en utilisant la navigation par planification, et active $RechargeE_{Pot}$ lorsqu'il en est suffisamment près.

Lors de la seconde série d'essais, la ressource d'E_{Pot} située en bas de l'environnement est ajoutée, de sorte qu'elle ne figure pas dans la carte utilisée par le robot, mais qu'elle apparaît lorsqu'il se dirige vers la ressource fixe. Le comportement du robot est alors variable selon le mode de calcul des saliences adopté (tab. 4.2).

Dans un premier temps, 15 essais ont été menés avec la configuration de calcul de saliences de l'expérience 1 (voir C.3). Cette configuration favorise la stratégie de navigation par planification vis-à-vis de celle par approche d'objet. En effet, dans le calcul des saliences de la boucle ventrale, le terme correspondant à cette première stratégie est pondéré par $0,65$ alors que la seconde n'est pondérée que par $0,55$. Ce réglage particulier est dérivé du fait que dans l'expérience 1, la planification était fondamentale pour résoudre la tâche. Il s'est avéré que pour ces 15 essais, le robot a suivi les indications de la planification 13 fois et ne s'est détourné de sa route de façon opportuniste que 2 fois (voir tab. 4.2).

Si l'on diminue la pondération de la planification à $0,55$, c'est-à-dire à la même valeur que celle de l'approche d'objets, le robot n'exhibe plus de politique nette concernant son choix (voir tab. 4.2). Arrivé à l'intersection, les saliences incitant le robot à descendre ou à continuer vers la gauche sont à peu près équivalentes, et ce sont les perturbations induites par le terme d'exploration qui déterminent le choix final.

Enfin, lorsque l'on donne l'avantage à la vision (pondération de la planification à $0,45$), le robot devient nettement opportuniste (voir tab. 4.2).

Evaluées par le test exact de Fisher, les différences entre les deux premiers ou les deux derniers cas ne sont pas significatives à 1% ($p = 0,05$ et $p = 0,11$), mais elles le sont en revanche entre le premier et le dernier ($p < 0.01$). Conséquemment, favoriser la planification ou l'approche

d'objets dans la pondération des calculs de salience génère des différences comportementales significatives. Il apparaît que le robot n'est pas systématiquement opportuniste, mais qu'il peut être réglé pour l'être, aussi bien que pour persister dans ses plans.

Discussion

Le fait qu'une pondération équivalente des deux stratégies conduise à une certaine forme d'indécision n'est pas totalement satisfaisant. En effet, se diriger vers une ressource visible proche plutôt que persister à rejoindre une ressource de même nature invisible plus éloignée semble être un choix systématiquement pertinent. La sélection opérée par la boucle ventrale concerne le *choix d'une direction de locomotion* à partir de saliences résultant de la combinaison de profils de direction issus de plusieurs stratégies de navigation. Il ne s'agit donc pas d'une *sélection des stratégies* elles-mêmes, ce qui explique peut être pourquoi cette expérience qui nécessite de favoriser explicitement une stratégie au regard d'une autre ne donne pas systématiquement des résultats tranchés. On pourrait par conséquent proposer une révision du modèle, inspirée de la proposition de Monchi *et al.* (2000), où le rôle de la boucle ventrale serait de sélectionner explicitement la stratégie à favoriser à un instant donné. Une fois cette stratégie sélectionnée, une boucle locomotrice dorsale serait en mesure de sélectionner la direction du déplacement à partir d'un profil de direction issu de cette seule stratégie.

Expérience 3 : disparition de ressources et oubli

L'expérience 3 est en quelque sorte l'opposé de la précédente. L'environnement de test est en effet le même, mais les deux ressources sont présentes lors de la phase d'apprentissage de la carte, et l'une d'elles est retirée lors des essais. Le robot est placé à la même position initiale et dans le même état interne initial que précédemment. La condition d'arrêt est identique. On s'intéresse ici au temps nécessaire pour que l'oubli de la ressource disparue soit suffisant pour que le robot s'intéresse à celle qui est toujours présente.

Résultats

Une série de 15 essais a été réalisée. Durant ces essais, le robot se dirige de façon systématique vers la ressource absente pour se recharger car elle figure toujours dans sa carte et s'avère être la plus proche de sa position de départ. Lorsque, s'approchant de la ressource absente, le robot se localise dans sa carte sur des nœuds associés à l'*Energie Potentielle*, le processus d'oubli se met en marche et le robot diminue progressivement la force de l'association entre ces nœuds et l'*Energie Potentielle*. Cette force diminuant, la planification prend de moins en moins compte de ces nœuds jusqu'à ce que la ressource fixe soit relativement plus attirante. Le robot se dirige alors vers elle et s'y recharge.

Le temps minimal nécessaire pour aller de la position de départ jusqu'à la ressource fixe

en faisant un crochet vers la ressource disparue, sans s'y arrêter, compte tenu des vitesses de translation et de rotation du robot, est de $46s$. Les durées des 15 essais ont une moyenne de $177,8s$ ($\sigma = 78,23$), soit $2min58s$, et ne dépassent pas $5min36s$. On constate donc que le temps d'oubli nécessaire pour modifier le comportement du robot, dans cette expérience, est compris entre 3 et $5min$. Cela semble relativement lent, compte tenu du fait que le temps de survie total avec une batterie pleine est de $33min$.

Cette lenteur s'explique par le fait que l'oubli au niveau d'un nœud est pondéré par la probabilité d'être présent en ce nœud (voir equ. 4.13). Or, à un instant donné, le positionnement du robot voit généralement sa probabilité répartie sur plusieurs nœuds (de l'ordre de la dizaine), la probabilité d'être présent en un nœud n'excédant alors rarement $0,25$, ce qui ralentit d'autant l'oubli.

On notera que pour engager l'approche de la seconde ressource, le robot n'a pas besoin d'oublier totalement la présence de la ressource disparue : il suffit que la force des nœuds concernés soit suffisamment faible. Il garde donc une trace de la ressource disparue.

Discussion

Cette troisième expérience montre que le processus d'oubli ajouté au modèle de navigation lors des modifications ayant permis son interfaçage avec la sélection de l'action, permet au robot de survivre dans un environnement où la disponibilité des ressources est variable. L'oubli considéré ici est relativement lent, mais il serait facile de l'accélérer en ajoutant un gain supérieur à 1 dans l'équation 4.13. Une discussion peut cependant être engagée sur la vitesse d'oubli souhaitable. En effet, est-il adaptatif de ne pas oublier trop vite un lieu susceptible de contenir à nouveau les mêmes ressources dans le futur ? Ne vaut-il au contraire mieux ne pas surcharger la carte cognitive avec des informations obsolètes ? Si tel est le cas, à quel niveau une accélération du processus doit-elle être gérée ? Les réponses à ces questions sont susceptibles d'êtres variables d'un environnement à l'autre, selon que les ressources y sont pérennes ou non, selon qu'elles apparaissent périodiquement au même endroit ou non, etc. La structure nerveuse codant les associations entre lieux et ressources n'est pas encore clairement définie, l'éventuelle prise en compte des récompenses dans les cellules de lieu de l'hippocampe et sujette à débat (Holscher *et al.*, 2003; Kobayashi *et al.*, 2003; Tabuchi *et al.*, 2003). Une étude plus approfondie des structures nerveuses en charge de cet apprentissage permettrait de remplacer la solution ingénieur ajoutée au modèle de navigation initial pour la remplacer par un modèle biomimétique plus adaptatif.

Expérience 4 : zone dangereuse

L'expérience 4 est menée dans l'environnement de la figure 4.16. Il mesure $10 \times 6m$, comporte deux ressources d'*Energie Potentielle* et une *Zone Dangereuse*. Le robot est initialement

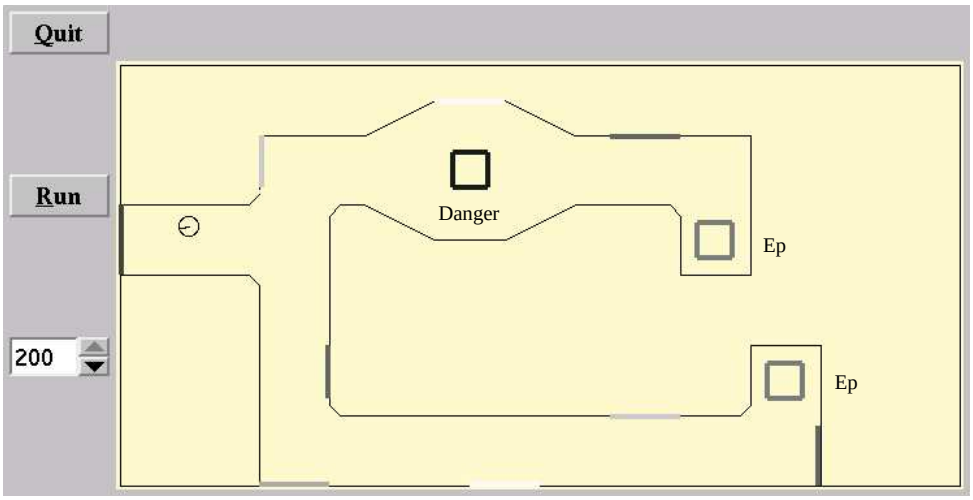

FIG. *4.16: Expérience 4 : environnement de test comportant deux ressources d'*Energie Potentielle *et une source de danger. Le robot est positionné à l'emplacement initial commun à tous les essais. La source d'*Energie Potentielle *la plus proche de la position de départ du robot est dangereuse d'accès, alors que la plus éloignée est sûre.*

positionné dans une niche (voir fig. 4.16). Il est en situation de manque d'*Energie Potentielle* plus ou moins fort ($E = 1$ et $E_{Pot} = 0,1$ ou $E_{Pot} = 0,5$), et sa *Peur* est constante, fixée à $0,2$. La ressource d'*Energie Potentielle* la plus proche du robot (environ $7m$) est située derrière la *Zone Dangereuse*, alors que celle qui est plus éloignée (environ $10m$) est sûre.

Le calcul des saliences de la boucle ventrale est modifié de façon à intégrer le profil de direction $\mathbf{Prox(31)}$ qui marque la proximité d'une *Zone Dangereuse*. Il est utilisé de sorte à défavoriser les déplacements menant vers les *Zones Dangereuses* visibles (voir tab C.6 en annexe).

Le calcul de $\mathbf{Plan}$ est également modifié : un profil de direction $\mathbf{P(31)}$ est également élaboré pour les ressources de type *Zones Dangereuses*, cependant il exerce un effet inhibiteur sur $\mathbf{Plan}$. Le calcul des composantes de $\mathbf{Plan}$ devient alors le suivant :

$$\begin{aligned}\mathbf{Plan}_i = \; & 1 - [(1 - motiv(E_{Pot}) \times \mathbf{P(127)}_i) \times (1 - motiv(E) \times \mathbf{P(255)}_i)] \\ & - motiv(ZD) \times \mathbf{P(31)_i}\end{aligned}$$

En l'absence de *Zone Dangereuse*, la ressource d'*Energie Potentielle* la plus proche génère dans le profil de direction issu de la planification des activations plus fortes que la ressource éloignée. Lorsque le robot arrive au point de jonction, il choisit donc systématiquement de se diriger vers la ressource proche. L'introduction d'une *Zone Dangereuse* située dans la même direction que la ressource d'*Energie Potentielle* proche génère une inhibition du profil de direction de planification dans cette direction. Si la ressource proche ne génère pas une activation

Etat interne		Choix	
Peur	$\mathbf{E_{Pot}}$	**Danger**	**Détour**
0,2	0,1	13	7
0,2	0,5	2	18
Test de Fisher		**p<**	0.01

TAB. *4.3: Expérience 4 : décompte des choix opérés selon l'état interne du robot (besoin en E_{Pot} plus ou moins urgent).*

suffisante pour contrebalancer cette inhibition, le robot choisira de rallier la ressource éloignée mais sûre.

Résultats

Deux séries de 20 essais ont été réalisées, alors que le robot avait un besoin modéré ($E_{Pot} = 0,5$) ou important ($E_{Pot} = 0,1$) d'*Energie Potentielle*. Dans le premier cas, le robot rejoint préférentiellement la ressource éloignée (voir tab. 4.3). En effet, l'attirance générée par la ressource proche, qui dépend directement du niveau d'*Energie Potentielle*, est atténuée par l'effet inhibiteur de la *Zone Dangereuse*. Au contraire, dans le second cas, il prend le risque de traverser la *Zone Dangereuse* pour rejoindre la ressource proche (voir tab. 4.3). En effet le besoin d'*Energie Potentielle* est tel que l'attirance générée par cette ressource proche est suffisante pour surpasser celle de la source éloignée, malgré l'inhibition. Les différences comportementales entre ces deux situations sont significatives (tab. 4.3).

On notera cependant que les comportements observés ici sont spécifiques d'une configuration environnementale particulière. D'une part, si l'on éloigne encore la ressource la plus éloignée, le robot choisit systématiquement la ressource proche. En effet, l'atténuation due à la distance est alors telle que la source proche, même atténuée, est toujours la plus attirante. De la même façon, si l'on fixe un niveau de *Peur* plus élevé, l'atténuation est telle que la source proche est toujours ignorée. D'autre part, si l'on modifie la configuration de sorte à n'avoir qu'une seule ressource d'*Energie Potentielle* accessible par deux chemins (voir fig. 4.17), il est possible que les effets de la *Zone Dangereuse* non seulement incitent le robot à emprunter le chemin le plus long, mais l'empèchent également de rejoindre la ressource en fin de parcours.

Discussion

Cette quatrième expérience montre que l'influence des variables internes sur le système de planification permet d'obtenir dans une même configuration spatiale des comportements variés, adaptés à l'état interne du robot. En effet, selon son degré de détresse énergétique, le robot est

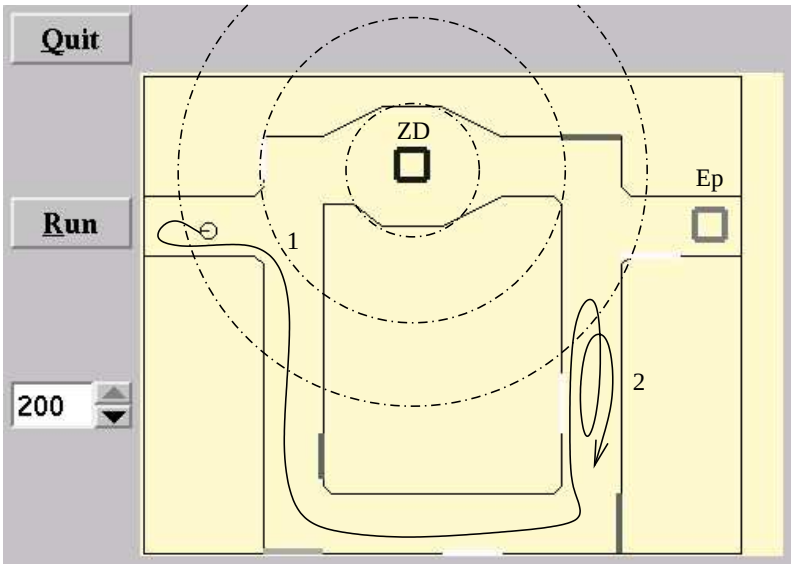

FIG. *4.17: Exemple d'environnement où l'influence* inhibitrice *de la* Zone Dangereuse *est telle que le robot choisit non seulement de prendre le chemin le plus long, mais ne peut rejoindre la ressource d'*Energie Potentielle. *En 1, le robot est attiré par E_P à la fois par le Sud et le Nord-Est, cependant, l'inhibition exercée par ZD sur les directions pointant vers de Nord-Est en cet endroit font que le robot est plus attiré par le Sud. En 2, le robot est attiré vers le Nord par E_P, mais lorsqu'il s'approche trop de ZD, cette attirance est totalement inhibée, il ne peut donc poursuivre vers le Nord.*

capable de prendre ou non des risques.

Les difficultés évoquées par la figure 4.17 sont dues à l'approche « plans-as-resources » adoptée. En effet, lorsque le robot est confronté à un problème de planification (par exemple en position 1 sur la fig. 4.17), le processus de planification que nous avons utilisé ne détermine pas une séquence globale de mouvements représentant la trajectoire permettant de rejoindre le but. Il génère au contraire une politique locale permettant, en chaque lieu de la carte, de savoir dans quelle direction aller pour rejoindre le but. Grâce à la production de cette politique locale en chaque lieu, il est possible de s'écarter temporairement du plan initial si un module autre que la planification le suggère (par exemple pour éviter un obstacle imprévu ou effectuer un détour opportuniste vers une ressource nouvelle), puis de suivre à nouveau les indications de la planification à partir de la nouvelle position. Il est donc possible de ne pas donner un contrôle absolu du robot à la planification et donc de fonctionner sur un mode « plans-as-resources ». Cet avantage est contrebalancé par les inconvénients rencontrés, semblables aux minimums locaux des approches de navigation par champ de force (Korenzy et Borenstein, 1991). Il semble donc qu'il faille à terme remplacer la méthode de planification locale adoptée par une méthode plus globale –planifiant une trajectoire dans son ensemble plutôt que des directions à suivre localement– tout en essayant de conserver le choix de l'approche « plans-as-resources ».

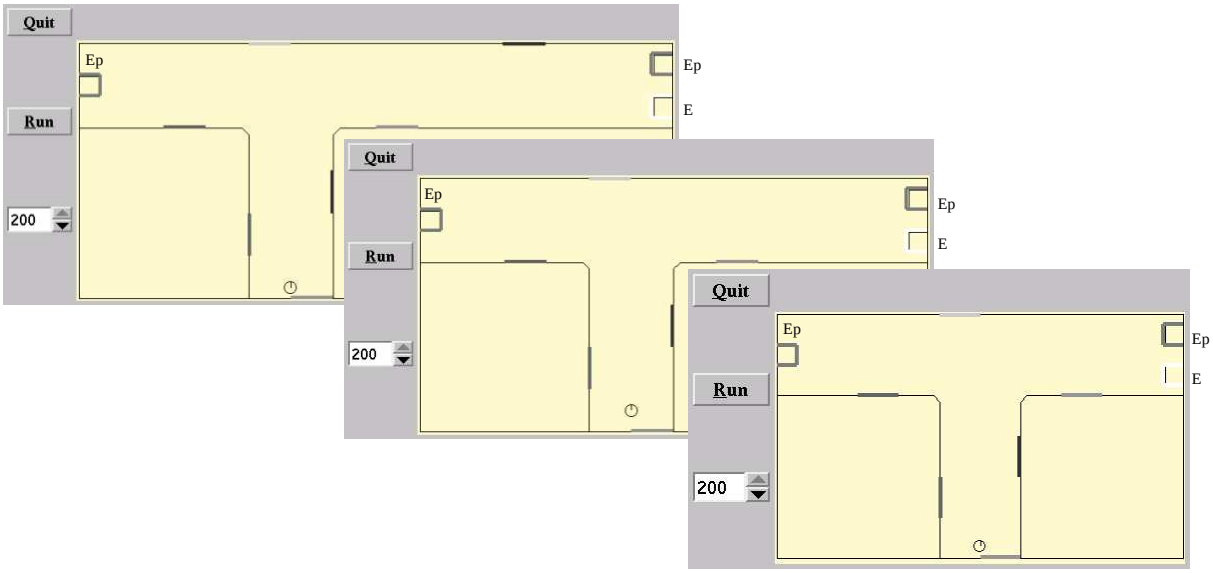

FIG. *4.18: Expérience 5 : labyrinthe en T, dont une branche ne contient qu'une ressource d'*Energie Potentielle *et l'autre une ressource d'*Energie Potentielle *et une ressource d'*Energie. *La longueur de la branche à deux ressources varie d'une expérience à l'autre. Le robot est positionné à l'emplacement initial commun à tous les essais.*

Expérience 5 : proximité de différents types de ressources

L'expérience 5 est inspirée d'une expérience réalisée dans (Quoy *et al.*, 2002) afin d'étudier les comportements générés par le couplage de deux motivations. Dans cette expérience initiale, le métabolisme du robot nécessite qu'il visite régulièrement trois type de ressources (nid, nourriture et boisson). La branche principale du T contient le nid, la branche gauche contient une source de nourriture alors que celle de droite contient de la nourriture et de la boisson. Cette branche contenant deux ressources peut être allongée jusqu'à deux fois la longueur de la branche ne contenant que de la nourriture.

Le comportement jugé souhaitable dans cette situation est celui qui consiste à emprunter préférentiellement la branche menant à l'eau et à la nourriture en même temps plutôt que celle ne contenant que de la nourriture. En effet, si les besoins en eau et nourriture augmentent à la même vitesse, il est moins coûteux de parcourir la branche longue de l'environnement pour atteindre les deux ressources que de faire un détour pour atteindre la nourriture proche, sachant qu'il faudra de toute façon emprunter la branche longue pour atteindre l'eau après.

L'expérience mise en œuvre ici diffère quelque peu de cette expérience originale. En effet, le métabolisme que nous avons utilisé ne dépend que de deux types de ressources. Par conséquent, notre protocole est le suivant : le robot est placé dans la branche médiane d'un labyrinthe en T (fig. 4.18), dont la branche gauche contient une ressource d'E_{Pot} et celle de droite une ressource d'E_{Pot} et une ressource d'E. Il est alors à la fois en manque d'*Energie* et d'*Energie Potentielle*

Rapport	G	D
1	3	12
1,5	4	11
2	8	7

TAB. *4.4: Expérience 5 : décompte des choix opérés selon le rapport entre la longueur de la branche droite et celle de la branche gauche. D : droite ; G : Gauche.*

($E = 0,5$ et $E_{Pot} = 0,5$). Le critère d'arrêt est le déclenchement de $RechargeE_{Pot}$.

Les premiers essais dans cet environnement ont été menés avec les calculs de salience utilisés dans les expériences 1 à 3 (voir tab. C.4, en annexe). Mais, ainsi que cela a été évoqué dans l'expérience 1, ces calculs sont incorrectement ajustés, de sorte qu'ils sont susceptibles de générer un blocage du robot à proximité d'une ressource dont il a grand besoin. Compte tenu du protocole expérimental, cette situation s'est systématiquement produite. Il est apparu que ce blocage n'est pas dû qu'à une trop forte « anticipation » d'une ressource visible. En effet, il semble également que le GPR soit en surcharge lorsqu'un trop grand nombre de ses canaux ont des saliences très élevées : il n'est plus capable de désinhiber suffisamment un canal pour générer un mouvement. Ceci se produit à l'arrivée à proximité d'une ressource d'*Energie Potentielle* déjà rencontrée, lorsque le robot a grand besoin d'*Energie Potentielle*, puisqu'alors, la détection visuelle et la planification envoient de très fortes activations sur de nombreuses directions contiguës. Afin de pallier ce problème, en plus d'une diminution des poids de la boucle dorsale, nous avons modulé l'influence de la planification par la présence d'une ressource visible, ce qui fait qu'implicitement, la planification a un rôle prédominant lorsque les ressources sont hors de vue, et qu'elle est remplacée progressivement par l'approche d'objet à mesure que les ressources sont de plus en plus visibles (voir tab. C.7, en annexe).

Résultats

Quinze essais ont été menés pour trois configurations du T : le rapport de longueur des deux branches y vaut 1, $1,5$ et 2. Du fait de son état interne initial, le robot a besoin des deux types de ressources. Par conséquent, tant que la branche de droite n'est pas beaucoup plus longue que celle de gauche (rapports de longueur de 1 et $1,5$), l'activité cumulée générée par les deux ressources de droite dépasse celle de la ressource de gauche (tab. 4.4). Cependant, lorsque les ressources commencent à s'éloigner, cette activité est atténuée par la distance et les deux options commencent à se valoir. De fait, dans nos séries d'essais, il apparaît qu'à partir du moment où la branche droite est deux fois plus longue que celle de gauche, le choix de la droite n'est plus systématique (voir tab. 4.4).

Discussion

Compte tenu du type de comportement attendu dans ce type d'expérience (se diriger systématiquement vers la branche droite), les résultats obtenus ne sont pas satisfaisants. Cependant, aucun mécanisme de notre modèle ne permet de compenser cette atténuation par la distance de l'attirance générée par les deux ressources. Le comportement obtenu n'a donc rien d'anormal compte tenu de l'état actuel de notre modèle.

Le modèle développé par Gaussier *et al.*, lui, intégre la notion de chemin « habituel » dans sa carte cognitive : les arêtes entre les nœuds régulièrement visités sont renforcées, de sorte que ces nœuds apparaissent moins distants qu'ils ne le sont en réalité. Cette notion de préférence permet de délaisser implicitement les trajectoires passant par des nœuds moins souvent visités. Cette particularité fait que le robot va délaisser la branche ne contenant que la nourriture et n'utiliser plus que la branche contenant les deux autres ressources d'autre part. En effet, cette branche contient deux ressources nécessaires à la survie, elle est nécessairement plus souvent visitée en début d'expérience que celle qui n'en contient qu'une. Le mécanisme de préférence va donc petit à petit la faire paraître moins longue qu'elle n'est en réalité, jusqu'à ce que la branche courte ne contenant que la nourriture ne présente plus d'avantage. Cette caractéristique est conservée même si l'on augmente la taille de la branche de droite jusqu'à deux fois la taille de celle de gauche.

Cette cinquième expérience montre que notre modèle nécessiterait d'être modifié pour mieux gérer ce problème, car il semble en effet que se diriger vers un lieu où deux ressources nécessaires sont réunies est plus adaptatif que de perdre de l'énergie à chercher l'une d'entre elles dans un autre lieu.

On peut se demander si la notion de chemin préférentiel intégré directement dans la carte cognitive représente dans tous les cas de figure un avantage adaptatif. La recherche de solutions en terme d'apprentissage de séquences comportementales relevant des comportements « habitudes » serait plutôt à envisager. Ce point sera repris en discussion générale.

Expérience 6

L'expérience 6 est une expérience de survie menée dans un environnement plus grand ($15 \times 15m$) et plus complexe que celui de l'expérience 1, où les situations-type testées jusqu'ici sont toutes susceptible de se produire (voir fig. 4.19 pour le détail du nombre de sources et leurs emplacements). En effet, tout d'abord, en dehors des sources d'E_{Pot} n°4 et d'E n°3, les sources sont hors de vue les unes des autres (Exp. 1). Ensuite, la source d'E n°1 diparaît (sa teinte passe de 255 à 245) et réapparaît toutes les demi-heures (Exp 2. et 3.). Deux chemins mènent de la source d'E n°1 à la source d'E_{Pot} n°1, le plus court passant à proximité de la *Zone Dangereuse*

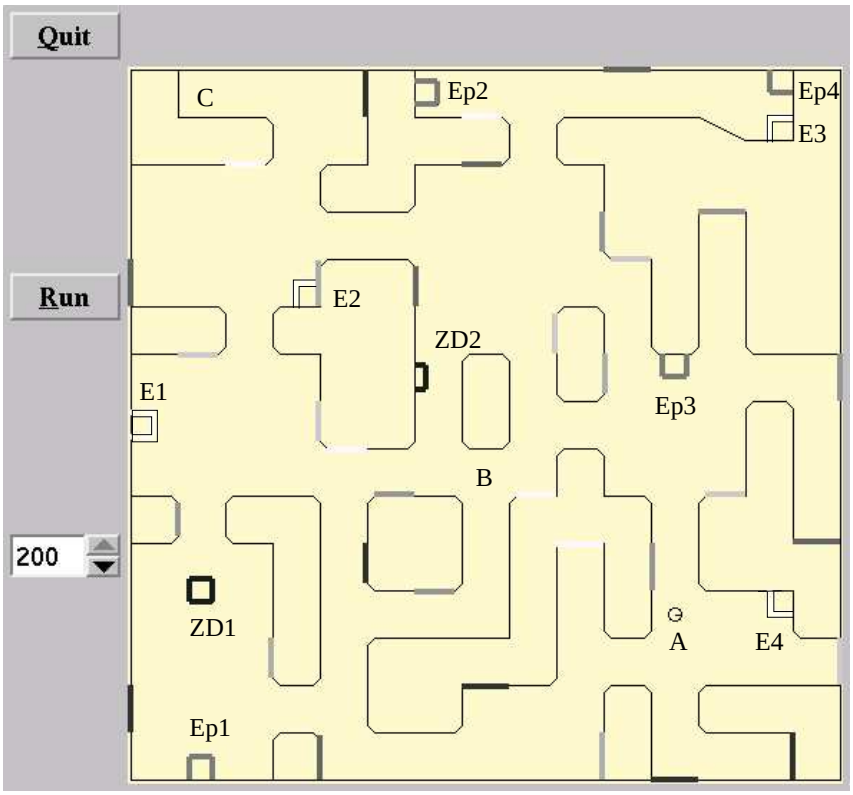

FIG. *4.19: Expérience 6 : Environnement complexe où sont susceptibles de se reproduire les situations-type précédemment testée. Il comporte 4 sources d'*Energie*, 4 sources d'*Energie Potentielle *et deux* Zones dangereuses*. A, B et C : emplacements de départ du robot.*

n°1 (Exp 4.). Enfin, les sources d'E n°3 et d'E_{Pot} n°2 et 4 sont dans une configuration similaire à celle de l'Exp. 5.

Le robot est initialement totalement chargé ($E = 1$ et $E_{Pot} = 1$), et bien localisé ($Des = 0$). Il n'a pas connaissance préalable de son environnement : sa carte cognitive est initialement vide.

Ainsi que cela a été fait pour l'expérience 1, nous nous sommes intéressés à la durée totale des essais menés dans cet environnement par un animat doté de la capacité de planification et par un animat réactif. Cependant, comme les essais n'ont pas de durée limitée afin que les animats aient la possibilité de parcourir l'environnement dans son ensemble, pour l'instant seuls six essais ont été menés pour la première configuration et quatre pour la deuxième, ce qui ne permet pas de comparaison statistique fiable (tab. 4.5). Pour ces essais, l'animat a été placé en trois positions différentes (A, B ou C, voir fig. 4.19) afin d'analyser l'effet plus ou moins favorable du contexte environnemental de la position initiale.

Résultats

Ces résultats préliminaires semblent montrer que le robot réactif, placé en (A) (fig. 4.19),

Planif.		**Réactif**	
Site	Durée (s)	Site	Durée (s)
A	75352	A	3345
A	64765	A	9079
B	3921	A	9344
B	12864	A	11064
B	2365		
C	24393		

TAB. *4.5: Expérience 6 : Emplacements de départ et durées (en s) des essais menés avec et sans planification.*

n'est pas capable de survivre beaucoup plus que $11000s$ (soit un peu plus de trois heures) alors que l'animat dont le système de navigation topologique, placé dans cette même position initiale, peut survivre jusqu'à près de $21h$. Ce dernier souffre cependant de la même limitation que dans l'expérience 1 : s'il ne parvient pas au début à la fois à cartographier sa zone de départ et à trouver un emplacement de recharge pour chacune des deux ressources (ce qui arrive fréquemment lors du départ en (B)), il meurt rapidement (ici deux essais ayant duré moins de $4000s$). En revanche, dès que cette condition est vérifiée, il devient capable de survivre très longuement (jusqu'à pratiquement $21h$).

L'examen des zones visitées par l'animat doté du système de navigation lors des quatre essais de longue durée (fig. 4.20 et 4.21) révèle certaines spécificités de son comportement.

Tout d'abord, il apparaît très nettement que, dans les quatre cas, l'animat explore une assez large part de l'environnement (plus des trois-quarts des zones de $1 \times 1m$ ont été visitées au moins une fois). Cependant, dès qu'il trouve une ressource de chaque type, il favorise les déplacements permettant d'aller de l'une à l'autre alternativement et semble limiter ses explorations aux seuls alentours immédiats de ces ressources.

Ensuite, on notera que dans le dernier essai (fig. 4.21, droite), l'animat a été d'une part confronté à la situation d'oubli de l'expérience 3 et d'autre part à celle de l'évitement d'une *Zone Dangereuse* de l'expérience 4. En effet, la source d'E n°1 disparaissant régulièrement, on constate que l'animat a été capable d'explorer et de trouver la source d'E n°2, qu'il a utilisé régulièrement. L'occupation de l'espace permet également de constater que pour rejoindre la source d'E_{Pot} n° 1, l'animat a employé à la fois le chemin dangereux (CH1) et le chemin long et sûr (CH2). Cependant, il est difficile de vérifier si ces choix ont été menés dans les conditions motivationnelles attendues, même si les résultats de l'expérience 4 incitent à le penser.

Enfin, dans les deux derniers essais, l'animat s'est manifestement trouvé bloqué longuement

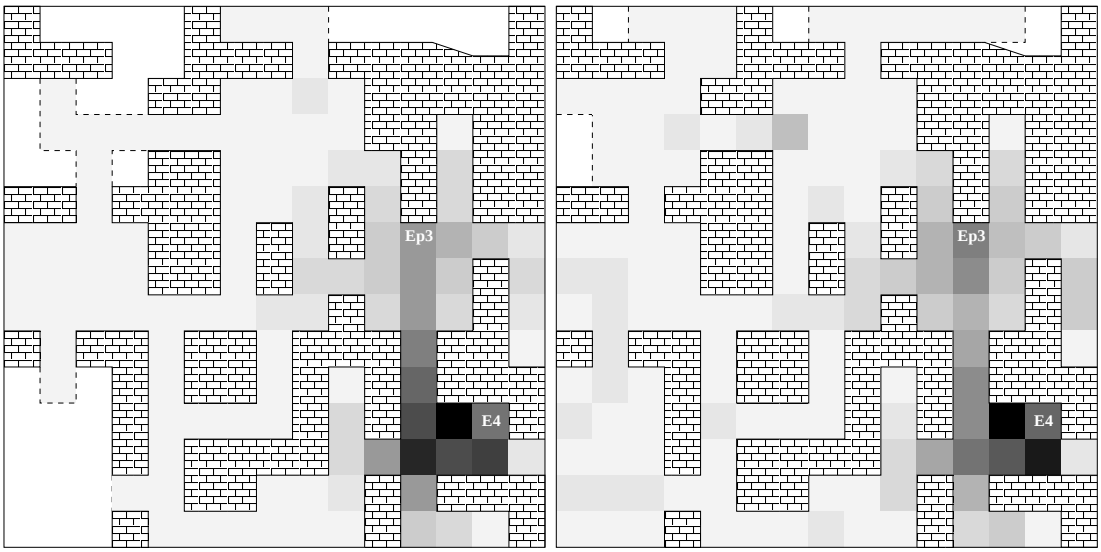

FIG. *4.20: Expérience 6 : occupation de l'espace durant les deux essais où l'animat débute au site A. Il assure sa survie en n'effectuant presqu'exclusivement des allers-retours entre* $E_{Pot}3$ *et* $E4$*. En pointillé : limite des zones où l'animat n'est jamais allé.*

dans des zones de l'environnement dénuées d'intérêt, c'est à dire dépourvues de ressources et n'étant pas sur un chemin menant d'une ressource à une autre (points A, fig. 4.21). Ce comportement est pour l'instant inexpliqué du fait de la difficulté à analyser dans le détail de nombreuses heures d'expérimentation. On peut cependant supposer que cela est lié, d'une part, à d'éventuelles surcharges du GPR qui immobilisent le robot et, d'autre part, à des erreurs de construction de la carte qui amènent la planification à proposer des directions de déplacement erronées, pointant dans des culs-de-sac.

Discussion

Les résultats préliminaires de cette sixième expérience semblent indiquer qu'un animat doté de notre modèle complet conserve sa capacité à survivre plus longtemps qu'un animat réactif, dans un environnement plus grand et plus complexe que celui de l'expérience 1.

Le fait de l'avoir situé dans un environnement plus complexe a également permis de révéler des limitations de ce système. Les essais correspondants nous renseignent sur l'impact très important que peut avoir le contexte environnemental sur les comportements du robot, de par la sensibilité observée de son temps de survie à ses positions initiales. En dehors des effets du hasard, susceptibles de l'empêcher de trouver les deux types de sources avant de mourir, il semblerait que notre modèle souffre de faiblesses encore non identifiées qui causent des blocages dans des zones inintéressantes de l'environnement. Les erreurs de cartographie et les surcharges du GPR, évoquées précédemment, y participent probablement. Une inadaptation des calculs de salience ajustés à la main peut également être la cause de l'occasionnel manque de pertinence

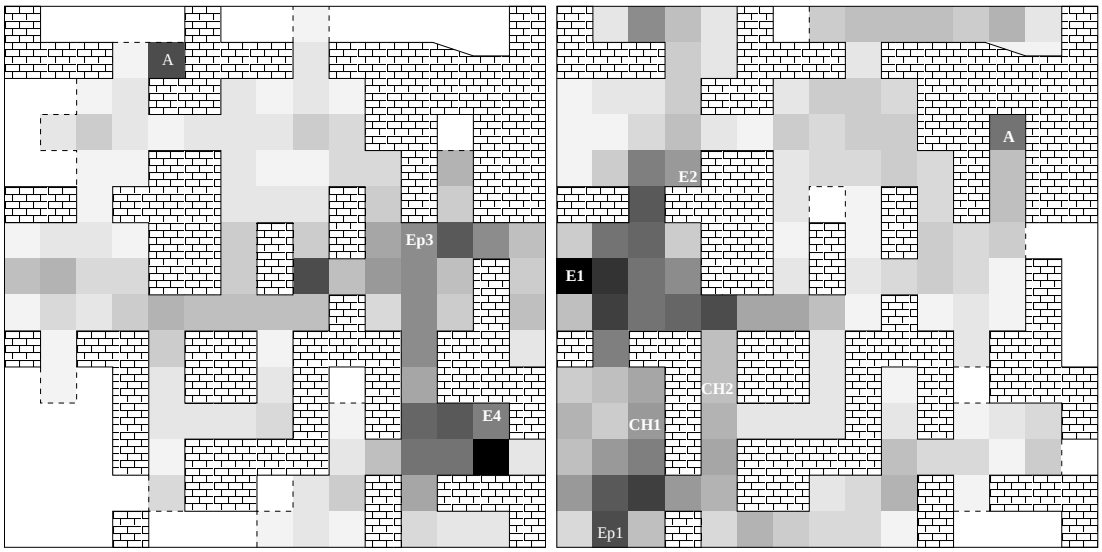

FIG. *4.21: Expérience 6 : occupation de l'espace durant (gauche) l'essai long ayant débuté en B et (droite) celui ayant débuté en C. L'animat assure sa survie en n'effectuant presqu'exclusivement des allers-retours entre* $E_{Pot}3$ *et* $E4$ *(gauche) ou* $E_{Pot}1$, $E1$ *et* $E2$ *(droite). Il semble que l'animat soit resté bloqué dans des certains lieux de l'environnement (marqués A), pour des raisons non-identifiées. Gauche : la présence d'une zone dangereuse entre* $E1$ *et* $E_{Pot}1$ *fait que l'animat passe de l'une à l'autre par les chemins* $CH1$ *ou* $CH2$ *selon son état motivationnel. En pointillé : limite des zones où l'animat n'est jamais allé.*

comportementale.

L'intégration de mécanismes d'apprentissage permettant de renforcer de bons enchaînements comportementaux et d'abandonner ceux qui s'avèrent inefficaces se révèle indispensable. De nombreux modèles computationnels déjà décrits (cf. 2.2.1) ainsi que de récents travaux de neurobiologie, que nous évoquerons dans la discussion générale du chapitre suivant, offrent des pistes pour leur modélisation.

Les difficultés méthodologiques concernant l'évaluation de la sélection de l'action, évoquées au chapitre 1, sont ici flagrantes.

D'une part, on vient de constater que la construction d'un environnement de test particulier peut avoir des conséquences importantes sur les résultats obtenus. Grâce à la configuration spécifique utilisée ici, nous avons pu mettre en évidence la variabilité du temps de survie de notre animat. Une autre configuration aurait pu lui donner un avantage considérable et masquer les améliorations nécessaires au modèle. Une méthodologie consistant à classer les environnements d'après leurs difficultés, recommandée depuis longtemps par Wilson (1991), serait à envisager.

D'autre part, pour une évaluation de l'adaptivité de l'animat, les sorties comportementales doivent pouvoir être visualisées de manière efficace. Cependant, lorsque les expérimentations

atteignent plusieurs heures –voire dizaines d'heures– de fonctionnement continu, il devient difficile d'analyser localement le comportement de l'animat pour savoir à quels moments précis il a effectué ou non des sélections pertinentes. Seuls des critères très globaux (durée de survie, occupation de l'espace) sont suffisament parlants, mais manifestement pas suffisamment précis pour analyser parfaitement le comportement du robot, afin, entre autres, de l'améliorer.

C'est pourquoi les différentes expériences réalisées dans ce chapitre sont complémentaires. Des environnements contrôlés, où le comportement efficace du robot peut être connu a priori, permettent d'évaluer quantitativement son adaptativité. Un environnement plus « naturel » peut révéler des enchaînements comportementaux non intuitifs qui auraient été occultés dans les précédentes expériences.

Au terme des expériences du chapitre 3, nous avions fait le point sur les fonctionnalités que pouvait apporter à Psikharpax le GPR sans système de navigation. L'ensemble des expériences menées dans ce chapitre montrent que notre modèle d'interface de la navigation et de la sélection de l'action permettrait de le doter d'un certain nombre de fonctionnalités supplémentaires à savoir (voir l'introduction pour la liste complète des fonctionnalités attendues) :

(ii) utiliser des stratégies efficaces d'exploration de l'environnement et de détection des amers rencontrés,

(iii) fusionner les informations visuelles acquises sur ces amers avec les informations proprioceptives concommitantes, afin de pondérer leurs influences respectives en fonction du contexte et d'élaborer une « carte cognitive » de son environnement,

(iv) utiliser cette carte pour se positionner lui-même et pour localiser les endroits où des récompenses ou des punitions ont été reçues,

(vi) choisir la stratégie de navigation la plus adaptée pour rejoindre un lieu où le but courant peut être satisfait, selon que ce but est directement visible ou qu'il est mémorisé dans la carte cognitive

Les diverses contributions et limitations, ainsi que les perspectives de ce travail, sont listées au chapitre suivant.

Chapitre 5

Discussion et Perspectives

[parlant des modèles computationnels :]
biologically not so inaccurate
Peter Redgrave (2003)

5.1 Contributions

5.1.1 Adaptation du GPR à une tâche de survie

Dans la première série d'expérimentations concernant la sélection de l'action, nous avons montré que le modèle de Gurney, Prescott et Redgrave, qui est le modèle des ganglions de la base le plus détaillé parmi ceux existant actuellement, est apte à résoudre une expérience de survie minimale sur un robot.

Par l'utilisation d'un mécanisme de sélection de type « winner-takes-all » appliqué à des canaux ségrégés correspondant à des actions différentes, complété d'une boucle de contrôle et d'une boucle de rétroaction modulant cette sélection, il permet une sélection de l'action par désinhibition sélective capable d'effectuer des choix efficaces malgré les imprécisions, tant au niveau sensoriel que moteur, générées par l'emploi d'un robot réel.

Les modifications que nous avons apportées au calcul des saliences (ajout de neurones sigma-pi et de fonctions de transfert) ont permis la résolution d'une tâche de survie en prenant en compte conjointement les incitations à agir provenant des variables internes et externes.

Dans l'environnement de test simple que nous avons utilisé, les effets dynamiques, induits par les boucles qu'il comporte, ne lui ont pas procuré de grand avantage en terme de survie vis-à-vis d'un simple WTA. En revanche, ils permettent l'utilisation de stratégies comportementales différentes d'un WTA, conséquences d'une persistance comportementale, qui se sont avérées plus adaptatives localement.

Le maintien des variables internes dans une meilleure *zone de confort* autorise en effet

l'animat à ne pas consacrer tout son temps à agir pour sa survie, mais lui donne la possibilité d'effectuer d'autres tâches.

Les oscillations comportementales, pouvant être induites ou limitées, lui permettent d'ajuster son comportement en fonction de contextes variés. Si l'animat doit éviter un « prédateur » pendant qu'il se recharge, une oscillation entre « guetter » et « se recharger » peut s'avérer adaptée. En revanche, osciller entre deux directions opposées menant à deux ressources nécessaires peut être fatal.

Enfin, une meilleure exploitation de la capacité à prélever moins d'énergie dans l'environnement lui donne à l'évidence un avantage en cas de raréfaction des ressources.

5.1.2 Interfaçage de la navigation et de la sélection de l'action

Le modèle proposé d'interfaçage de la navigation et de la sélection de l'action est une extension du GPR. Deux circuits des ganglions de la base ont été modélisés, selon l'hypothèse qu'un circuit dorsal moteur permettrait de sélectionner les actions non locomotrices –ici les comportements de recharge– et qu'un circuit ventral issu du noyau accumbens « core » interviendrait dans la sélection de profils de direction d'actions locomotrices liées à la navigation.

Les interactions entre ces deux circuits ont été modélisées par des projections du circuit dorsal vers le circuit ventral au niveau du noyau subthalamique. Une autre hypothèse neurobiologique –l'interaction par voie cortico-corticale– a été modélisée, mais elle a conduit à une mauvaise coordination des comportements et des profils de direction sélectionnés.

Ce modèle contribue ainsi à la construction de modèles computationnels intégrant navigation et sélection de l'action, sans que l'une et l'autre ne soient ni confondues, ni négligées (cf 1.2.3). A notre connaissance, seul notre modèle prend en compte cette intégration en s'inspirant d'hypothèses émises par les travaux neurobiologiques concernant le rôle d'interface du nucleus accumbens.

Il contribue aussi à la construction de modèles computationnels gérant l'interaction de plusieurs stratégies de navigation (cf 4.2) –ici : exploration, approche d'objet et planification topologique– dont la nécessité avait notamment été constatée par Trullier (1998) et Arleo (2000) en conclusion de leurs thèses portant sur la modélisation de la navigation inspirée de l'hippocampe.

Le modèle a globalement donné satisfaction en terme d'efficacité pour le contrôle d'un robot simulé.

Il a en effet réussi, dans un certain nombre de situations-types, à opérer de la manière attendue. Tout d'abord, l'ajout de la navigation topologique au système de sélection de l'action

s'est avéré efficace, puisqu'un robot doté de cette nouvelle capacité s'est révélé capable de survivre dans une configuration environnementale mettant en difficulté un robot réactif. Ensuite, le modèle s'est avéré capable de s'adapter aux changements de l'environnement, d'une part en exhibant des capacités d'opportunisme lors de l'apparition de nouvelles ressources, grâce à l'interaction de deux stratégies de navigation (l'approche d'objets et la planification topologique), et d'autre part en se montrant capable d'oubli lors de la disparition de ressources connues. Enfin, il s'est vu exécuter des choix adaptés à la fois à son état interne et à la configuration de l'environnement dans une expérience d'évitement de zone dangereuse et, de manière plus limitée, dans la reproduction d'une expérience de labyrinthe en T où il s'agissait de choisir la branche contenant deux ressources vitales même si cette branche était la plus longue.

Enfin, il s'est montré capable de survivre longuement dans une expérience de survie dans un environnement de grande taille, à la structure complexe et susceptible de présenter les situations-types précédemment testées.

5.2 Comparaison avec le modèle de Gaussier *et al.*

Comme nous l'avons évoqué en 4.3.3, le modèle de Gaussier *et al.* (2000) est le seul actuellement qui pourrait fournir à un animat des capacités similaires à celles du modèle que nous avons proposé au chapitre 4. La comparaison avec le modèle de Guazzelli *et al.* nécessiterait qu'il soit doté d'un métabolisme et testé dans une tâche de survie, sans compter que la différence dans le niveau de modélisation adopté fausserait sans doute cette comparaison.

La comparaison avec l'architecture de Gaussier *et al.*, amorcée dans l'expérience 5 du chapitre 4, permet d'identifier les avantages et limitations respectives des deux modèles.

5.2.1 Modélisation biomimétique

Le modèle de Gaussier *et al.* s'intéresse à la formation hippocampique et à ses interactions avec le cortex préfrontal. Il s'arrête conséquemment au niveau du noyau accumbens, qu'il considère comme la couche locomotrice de sortie. Notre modèle aborde précisément la modélisation de l'intégration des informations issues de l'hippocampe et du cortex préfrontal dans le noyau accumbens et les circuits des ganglions de la base situés en aval, mais ne prétend pas utiliser un système de navigation topologique modélisant le cerveau du rat. Sur cet aspect biomimétique, les deux modèles sont donc complémentaires.

5.2.2 Planification

D'un point de vue fonctionnel, le modèle de Gaussier *et al.* implémente une navigation *topologique*. Ainsi que cela a été expliqué en 4.3.3, elle est fondée sur l'utilisation d'un graphe

de transitions entre lieux plutôt qu'un graphe de lieux. Les transitions menant à une ressource donnée sont associées au neurone de motivation correspondant par un apprentissage hebbien comparable à celui que nous avons utilisé pour l'association d'un lieu à une ressource. Enfin, la planification fonctionne par propagation de l'activation des neurones motivationnels dans le graphe. Elle est donc modulée à la fois par la distance aux ressources et aux forces des associations entre nœud du graphe et motivation ce qui donne au final un résultat équivalent à notre profil de direction **Plan**.

Ces similarités font que les deux modèles devraient exhiber des performances similaires dans le cadre des expériences 1 et 3 du chapitre 4, c'est-à-dire dans le cas de l'utilisation de la planification pour survivre dans un environnement où les ressources vitales sont éloignées, et dans celui de l'oubli dans la carte cognitive d'une ressource ayant disparu de l'environnement.

5.2.3 Chemins préférés et zones dangereuses

Dans le modèle de Gaussier *et al.*, les poids entre les nœuds de la carte cognitive ne représentent pas la distance exacte les séparant. En effet, non seulement ces poids diminuent au fur et à mesure que le robot emprunte les transitions correspondantes, de sorte que s'établissent des chemins « préférés », mais ils augmentent lorsque le robot traverse des zones dangereuses ou difficiles, de sorte que ces zones seront à l'avenir contournées, puisqu'un chemin les traversant semblera plus long.

L'expérience 5, inspirée de (Quoy et al., 2002), met l'animat en position de choisir entre une ressource vitale proche et deux ressources vitales éloignées. Elle a montré l'avantage adaptatif qu'est susceptible de fournir cette première notion de chemin « préféré ». Dans l'expérience originale, l'animat est capable de favoriser implicitement des choix intéressants à long terme en n'empruntant que la branche à deux ressources. Dans la nôtre, il ne l'emprunte que si la distance parcourue n'est pas trop longue, notre modèle étant intrinsèquement incapable d'exhiber un tel pocessus de préférence. Le traitement des zones dangereuses par accroissement apparent de la distance est plus simple que celui que nous avons adopté, qui consiste à inhiber les directions de déplacement rapprochant d'une zone dangereuse.

Cependant, cette introduction d'une distance apparente différente de la réalité, intégrée directement dans la carte cognitive, peut dans d'autres cas nuire à l'animat.

Par exemple, dans la configuration en T si, à un instant donné, le robot est chargé en *Energie* et n'a donc besoin que d'*Energie Potentielle*, il ne semble pas particulièrement adaptatif de choisir la branche la plus longue du T pour s'en procurer. On peut également imaginer qu'opter pour un chemin plus long en réalité que dans la carte, alors que le robot est en détresse énergétique, peut générer une mort inattendue, le chemin semblant assez court pour permettre la recharge avant le déchargement total.

De la même façon, l'allongement des chemins dangereux a l'inconvénient de ne pas pouvoir

proposer un comportement variable selon l'importance du besoin énergétique, permettant de choisir de prendre le risque de traverser la zone dangereuse lorsque le manque d'énergie est critique, comportement que notre modèle a exhibé dans l'expérience 4.

Il semble donc nécessaire de modifier notre modèle afin d'améliorer son comportement dans l'expérience du labyrinthe en T, mais en explorant des solutions computationnelles alternatives à celle de Gaussier *et al.* afin d'en éviter les inconvénients. Une des solutions envisagée est d'intégrer des processus d'apprentissages de séquences de comportements (cf. 5.5.2).

5.2.4 Fusion de stratégies de navigation

Le modèle de Gaussier *et al.* n'aborde pas la fusion de plusieurs stratégies de navigation. Or il s'avère qu'à portée visuelle des ressources, l'approche d'objets permet en général une approche plus précise et plus rapide que la planification topologique. Cette fusion entre stratégies aux propriétés complémentaires est une spécificité de notre modèle vis-à-vis de celui de Gaussier *et al.*.

5.3 Remise en question des options de modélisation

Les expériences menées dans ce travail ont mis en évidence un certain nombre de limitations du modèle de sélection de l'action et du modèle d'interfaçage de la navigation et de la sélection de l'action.

5.3.1 Capacités de sélection du GPR

Une limitation des capacités de sélection du GPR, qui n'avait pas été révélée dans les expériences de sélection de l'action seule (chapitre 3) est apparue lors de l'interfaçage de la sélection de l'action et de la navigation (expériences 1 puis 5 du chapitre 4)

Dans des situations où de nombreux canaux ont des saliences très importantes, leur désinhibition n'est pas suffisante pour passer le seuil d'activation. Des tests seraient nécessaires pour déterminer les conditions exactes dans lesquelles ce problème survient et si le remplacement des inhibitons latérales uniformes du striatum dans le GPR original par des inhibitions latérales graduelles (voir 4.4.4) y participe.

Le modèle GPR que nous avons utilisé a été sujet depuis à des modifications (Wood *et al.*, 2001; Humphries, 2002) tenant compte davantage des données biologiques :

– des résultats tendent à prouver que les neurones du striatum possédant des récepteurs dopaminergiques de type D1, se projetant sur l'EP, se projettent également vers le GP

(Wu *et al.*, 2000), l'intégration de ce fait dans le GPR favorise la sélection d'un seul canal à la fois,

– il a aussi été testé sans les inhibitions latérales du striatum, dont l'existence est sujette à controverse, et a conservé ses caractéristiques de sélection,
– enfin l'éventuelle présence d'inhibitions latérales dans l'EP et le GP a été testée et améliore légèrement les capacités de sélection du GPR.

Le remplacement du modèle GPR utilisé ici par sa dernière version, susceptible d'être plus efficace, est une étape logique de l'amélioration de notre modèle.

5.3.2 « Soft-switching » et « hard-switching »

Le modèle GPR développé à l'ABRG y a toujours été utilisé avec une sélection de type « soft-switching » en sortie. Les expériences menées s'y prêtaient très bien, aucun des comportements ne nécessitant un recrutement exclusif des effecteurs.

Cette approche du « soft-switching » n'est pas sans relation avec le modèle de sélection de l'action proposé par Tyrrell (1993a) où tous les comportements sont susceptibles de participer, à des degrés dépendant leur activité, à la sélection des actions locomotrices. On notera que ces seules les actions locomotrices étaient sujettes à ces éventuels compromis.

Nous avons adopté un positionnement comparable à celui de Tyrrell, puisque les actions locomotrices issues de la boucle ventrale sont issues d'un mécanisme de « soft-switching », alors que les actions de recharge issues de la boucle dorsale sont sélectionnées sur un mode « hard-switching ». Ce choix est issu du fait qu'il semble concevable de choisir de se déplacer dans une direction qui permet de se rapprocher de deux buts différents simultanément. Au contraire, amorcer partiellement et simultanément les actions permettant de boire et de manger ou essayer d'attrapper de la nourriture tout en s'en éloignant ne semble pas raisonnable. Il est manifeste que certains comportements nécessitent le contrôle exclusif des effecteurs et ne sont donc pas compatibles avec un mécanisme de « soft-switching ».

Il semble donc que d'un simple point de vue pratique, le « soft-switching » ne peut pas être utilisé de façon systématique, sans compter que des travaux plus théoriques tendent à montrer que l'usage de compromis dans la locomotion n'apporte pas nécessairement d'avantage en terme de survie (Crabbe, 2002).

5.3.3 Stratégies de navigation et sélection de l'action

Bien qu'intégrant deux stratégies de navigation, notre modèle peut être amélioré quant à la gestion de leurs interactions. Ainsi que cela a déjà été discuté dans le cadre de l'expérience 2 du chapitre 4 –au cours de laquelle l'animat ne choisit pas toujours l'approche d'objet pourtant plus efficace que la planification– il est possible que le choix d'une sélection de direction en boucle

ventrale en combinant des suggestions issues de plusieurs stratégies de navigation s'avère moins efficace que la sélection de la stratégie à appliquer à un instant donné. C'est une voie possible de modification du modèle, inspirée de (Monchi *et al.*, 2000), dont l'efficacité serait à comparer avec celle du modèle actuel. Une autre interprétation en terme de sélection du but pourrait également être explorée, afin de chercher à mieux rendre compte des résultats d'électrophysiologie concernant le « core » du noyau accumbens (Mulder *et al.*, submitted), évoqués en 4.1.1.

La question de la compétition directe entre stratégies de navigation pose de surcroît le problème de l'usage permanent de la cartographie et de la planification. En effet, la navigation *topologique* est complexe, coûteuse et peu précise. On imagine donc aisément que lorsque, par exemple, l'environnement est suffisamment connu pour que l'on puisse mettre en œuvre une *navigation par action associée à la reconnaissance d'un lieu* efficace, le système de navigation *topologique* pourrait être totalement désactivé pour n'être réactivé qu'à l'arrivée dans un environnement nouveau. Dans le cas d'une urgence énergétique, on peut également envisager de désactiver le système de navigation *topologique* au profit de l'exploration couplée avec l'approche d'objet, pour s'affranchir des précautions qu'il nécessite (retour sur ses pas régulier pour la constitution d'une carte robuste) afin d'explorer le plus rapidement possible l'environnement pour trouver la ressource nécessaire.

Le système de navigation que nous avons utilisé ne peut fonctionner correctement s'il est temporairement désactivé, la continuité des déplacements dans un même environnement étant nécessaire à une bonne localisation. La mise en œuvre d'une réelle compétition entre stratégies de navigation nécessitant de pouvoir le désactiver temporairement, il faudrait soit le modifier, soit en utiliser un autre capable de gérer plusieurs cartes. En effet, lorsque le système de navigation *topologique* est réactivé, s'il est incapable de reconnaître sa localisation et doit donc se considérer comme placé dans un nouvel environnement, il doit pouvoir construire une nouvelle carte indépendante des précédentes.

5.3.4 Modélisation et coordination des boucles

Seul le circuit issu du « core » du noyau accumbens, qui a un rôle dans l'intégration des données topologiques pour la locomotion, a été modélisé. Celui issu du « shell » semble jouer un rôle important en ce qui concerne les motivations (Kelley, 1999) et est en position de moduler l'activité de l'ensemble des autres circuits des ganglions de la base ainsi que du cortex via les projections dopaminergiques issues de l'aire ventrale tégmentale (Joel et Weiner, 2000). L'hypothèse qu'il effectuerait une pré-sélection des motivations, afin que l'ensemble des choix comportementaux à un instant donné soient influencés par la motivation la plus prioritaire, proche de la proposition de (Dayan, 2001), serait alors à tester. Ce choix d'une motivation prioritaire à un instant donné, en amont du choix comportemental permettant éventuellement

de la staisfaire, pourrait le simplifier et par là améliorer les performances. Toutefois, ce circuit « shell » n'aurait pas la structure d'un GPR (Thierry *et al.*, 2000) et un travail de modélisation supplémentaire serait à mener sur ce point.

Du point de vue des données biologiques disponibles, il apparaît que la question des voies employées pour la coordination entre boucles cortex-ganglions de la base thalamus-cortex n'est pas résolue. L'utilisation de la voie trans-subthalamique donne satisfaction d'un point de vue computationnel pour permettre à une boucle d'inhiber nettement et rapidement des décisions émanant d'une autre boucle. Cependant, le fait que les excitations provenant d'une région du noyau subthalamique dédiée à une boucle ne touche qu'une zone frontière, assez limitée, de la région de la substance noire réticulée de la boucle ciblée (Kolomiets *et al.*, 2003) tend à montrer que nous en avons fait un usage trop extensif. En effet, dans le modèle présenté, ces excitations touchent l'ensemble de la région de la substance noire réticulée cible. Des données expérimentales complémentaires sont nécessaires afin de mieux cerner comment ces interactions entre boucles fonctionnent et quels sont leurs rôles précis. L'enregistrement simultané de l'activité électrophysiologique de neurones du noyau accumbens et des régions motrices du striatum lors de la résolution par un rat de tâches nécessitant théoriquement la coordination de plusieurs boucles n'a pas encore été réalisé (Deniau, 2003) et serait un premier pas permettant de mieux comprendre les rôles respectifs des boucles et leurs interactions.

5.4 Validation des modèles

Ainsi que nous l'avons signalé en introduction, l'évaluation des modèles de sélection de l'action est problématique, nous y avons été confronté.

5.4.1 Choix des comparaisons

Pour réaliser cette évaluation, nous n'avons pu effectuer de comparaisons entre le fonctionnement de notre modèle et le comportement animal. Les comparaisons ont concerné, d'une part, le modèle GPR avec un mécanisme de sélection très simple (WTA) et, d'autre part, le modèle d'interface sélection de l'action/navigation avec ce même modèle privé des capacités de navigation topologique.

Les premières comparaisons ont permis de révéler des subtilités dans les stratégies comportementales du GPR, dont les effets relativement limités sont peut-être dûs à un environnement de test et à une tâche à résoudre trop simples. En effet, une sélection de l'action décentralisée permet aux animaux dépourvus de ganglions de la base de survivre efficacement. Les ganglions

de la base sont supposés répondre au besoin croissant de centralisation causé par la gestion d'un grand nombre de comportements complexes (Prescott, 2001).

Le second type de comparaison permet d'attester du bon fonctionnement du système de navigation topologique et de son interfaçage avec la sélection de l'action, mais également de montrer le réel avantage qu'il procure à un animat en terme de survie.

Du point de vue de la modélisation du vivant, nous regretterons qu'en l'état actuel, notre modèle ne soit pas en mesure de fournir des prédictions à tester en retour chez l'animal.

5.4.2 Evaluation du comportement

Du point de vue de la réalisation d'un mécanisme de sélection de l'action pour un robot autonome, malgré les réserves exprimées au chapitre 1, nous avons testé un certain nombre des « propriétés souhaitables » pour un tel système. Cependant, nous nous sommes efforcés de quantifier ces tests afin de ne pas se limiter à une simple constatation subjective.

En ce qui concerne les expériences de survie de longue durée, il apparaît qu'elles se prêtent très bien à l'analyse de critères très globaux (durée de vie, temps passé dans les diverses régions de l'environnement). Il est en revanche difficile de déterminer quelles autres mesures sont pertinentes pour analyser plus finement plusieurs heures d'expérimentations, et en particulier comment segmenter le comportement continu résultant pour dégager les éléments adaptatifs ainsi que les limitations.

5.4.3 Evaluation de l'« intelligence » du système

Il est souvent difficile de savoir si l'efficacité d'un mécanisme de sélection de l'action est due à l'habileté de l'expérimentateur à ajuster les paramètres de son système ou à une réelle amélioration conceptuelle dudit système vis-à-vis de ses prédecesseurs. Dans le cadre de nos expérimentations, il apparaît que des paramétrages particuliers permettent de résoudre ponctuellement des problèmes de sélection (voir l'expérience d'oscillations comportementales du chapitre 3 et l'ensemble des expériences du chapitre 4).

C'est pourquoi dans le chapitre 3 nous nous sommes efforcés de montrer les différences fondamentales induites par la présence d'une rétroaction positive dans le GPR (par exemple la possibilité d'ajuster la durée des oscillations comportementales alors que le WTA est condamné à osciller).

C'est également pourquoi dans le chapitre 4, après avoir été amené à modifier le calcul des saliences d'une expérience à l'autre pour satisfaire au mieux les exigences de chacune, nous nous sommes efforcés de fusionner l'ensemble de ces modifications pour proposer un calcul de saliences unifié, confronté à une expérience de survie en environnement complexe (voir 4.6.2.0, exp. 6).

5.4.4 Simulation vs. plateforme robotique

Les évaluations réalisées pour la seule sélection de l'action ont été menées à bord d'un robot réel, sans les biais qu'est suceptible d'introduire une simulation. Il s'agissait de la seconde implémentation robotique du GPR, pour la résolution d'une tâche et dans un robot différents, ce qui tend à montrer que son efficacité n'est pas conditionnée à des conditions particulières.

Au contraire, les évaluations du modèle d'interfaçage de la navigation et de la sélection de l'action ont été menées en simulation. Cette simulation a un degré de réalisme limité : d'une part, la caméra simulée à directement accès à la teinte de objets, sans être soumise aux problèmes d'éclairage variable auquel un robot réel sera forcément confronté, d'autre part, l'erreur faite sur l'estimation de la direction n'est pas cumulative au cours du temps, ce qui simule l'utilisation d'un compas. La résolution de ces problèmes sur un robot réel n'est cependant pas du ressort des ganglions de la base, la modélisation ne peut donc être remise en cause sur ce point.

Concernant l'intérêt du modèle pour une application sur robot réel, il semble que l'utilisation d'une carte de traitement visuel temps-réel (développée par BEV, partenaire du projet Psikharpax) et d'un compas magnétique ou du compas visuel développé par Gourichon *et al.* (2002) permettent d'envisager son portage sans que les simplifications de la simulations ne s'avèrent rédhibitoires.

5.5 Perspectives

5.5.1 Une navigation plus biomimétique

Comme nous l'avons indiqué en 4.5, le modèle proposé ici utilise un système de navigation qui n'est pas issu d'une modélisation biomimétique. La partie planification de ce modèle a été adaptée à nos besoins par une approche ingénieur, ce qui a abouti à des calculs assez complexes et bien éloignés d'une modélisation biomimétique.

A ce titre, ainsi que cela a déjà été évoqué, notre modèle semble complémentaire du travail de (Gaussier *et al.*, 2000). La combinaison de ces deux modèles semble donc une voie de développement possible, mais elle n'implémenterait que des stratégies de navigation déjà présentes dans notre modèle.

On peut donc également envisager de remplacer les processus de cartographie et de localisation par des modèles existants de l'hippocampe ((Arleo et Gerstner, 2000), par exemple). Ceux-ci mettant en général en œuvre des stratégies de *navigation par action associée à la reconnaissance d'un lieu*, cette stratégie de navigation supplémetaire enrichirait le répertoire de stratégies utilisable par l'animat. Cependant, ces modèles localisent en général l'apprentissage de cette stratégie au niveau des projections de l'hippocampe vers le noyau accumbens, ces pro-

jection ciblant préférentiellement la partie « shell ». Or, comme nous l'avons dit précédemment, l'apprentissage correspondant, de type S-R ou « habitudes », semble avoir lieu dans les boucles dorsales (Graybiel, 1998; Cardinal, 2001; Everitt et Wolf, 2002), ces modèles doivent donc probablement être modifiés sur ce point.

Enfin, ils doivent pour la plupart être dotés de capacités de planification, leur combinaison avec des méthodes de planification biomimétiques, inspirées des colonnes corticales (Burnod, 1989; Bieszczad, 1994; Frezza-Buet et Alexandre, 1999), est donc une voie supplémentaire à mettre en chantier. La façon dont sera modélisé ce futur modèle et ce qu'il pourra fournir en sortie déterminera sans doute de nouvelles hypothèses relatives à un interfaçage avec un modèle des ganglions de la base.

5.5.2 Apprentissage

Rappelons que notre travail a consisté à utiliser un modèle de sélection de l'action *sans mécanisme d'apprentissage*. Seul le mécanisme de navigation en était pourvu pour apprendre la localisation des ressources dans sa carte cognitive.

Les situations d'échec rencontrées dans l'expérience 1 du chapitre 4 montrent la difficulté qu'il peut y avoir à ajuster à la main le calcul des saliences de ce type de modèle : certains mauvais choix ne peuvent se révéler qu'assez tardivement. L'intégration de capacités d'adaptation en ligne (apprentissage par renforcement) semble être nécessaire à un tel système afin d'éviter les blocages dans des situations imprévues et non testées par le concepteur. Ce travail a été initié lors d'un stage de DEA dans notre laboratoire (Khamassi, 2003; Khamassi *et al.*, 2003).

Ainsi que nous l'avons évoqué plus haut, l'intégration de la statégie de navigation par action associée à la reconnaissance d'un lieu a été volontairement ignorée dans notre modélisation. Elle relèverait en effet tout d'abord de capacités d'apprentissage de comportements S-R dans les circuits dorsaux qui n'ont pas encore été intégrées au GPR. Les modèles d'apprentissage par renforcement présentés en 2.2.1 sont de bons candidats pour l'apprentissage de simples comportements S-R.

L'ajout de ces capacités dans une boucle dorsale motrice rendrait encore plus prégnant le problème des interactions entre boucles, puisque des décisions de déplacement, éventuellement contradictoires, seraient prises dans deux boucles distinctes des ganglions de la base et nécessiteraient donc la modélisation du mécanisme permettant de résoudre les conflits engendrés.

L'apprentissage permettant l'assemblage de ces comportements en séquences comportementales « habitudes » (Graybiel, 1998) est en revanche à explorer. En effet, les modèles traitant des séquences, présentés en 2.2.3, ne sont capable d'apprendre que des séquences explicitement

fournies par l'expérimentateur pour l'apprentissage. L'auto-analyse du comportement du système afin d'identifier de manière autonome les séquences d'actions fréquemment répétées, afin de créer de nouveaux comportements codant pour l'une de ces séquences entière est en dehors de leurs capacités. Un tel modèle fournirait pourtant une piste d'analyse du codage des cellules du noyau accumbens (voir 4.1.1), certaines semblant coder des séquences entières d'actions.

Ces deux problématiques relèvent de prochains travaux de thèse proposés conjointement par le LPPA et le LIP6.

Enfin, l'apprentissage permettant d'associer à un lieu stocké dans une carte cognitive à un certain type de récompense, modélisé par un simple apprentissage hebbien dans notre modèle, mériterait d'être approfondi. En effet, il implique probablement l'hypothalamus, mais la localisation exacte de ces associations, en particulier l'hypothèse qu'elles aient lieu dans l'hippocampe (Holscher *et al.*, 2003; Kobayashi *et al.*, 2003; Tabuchi *et al.*, 2003) reste à déterminer clairement.

5.5.3 Extension aux autres circuits des GB

Les rôles d'un certain nombre de circuits des ganglions de la base n'ont pas été explorés dans ce travail.

Pourtant, l'implication des ganglions de la base dans la mémoire à court terme et de travail (cf. 2.2.2) est particulièrement intéressante dans le cadre de la modélisation. En effet, les comportements sélectionnés par le GPR dans les expériences de Montes-Gonzalez (2000) ou dans nos propres travaux (Chap. 3) sont des automates à états finis nécessitant la mémorisation de l'état courant, dans ce que Montes-Gonzalez nomme d'ailleurs les STM pour « short-term memory ». Le remplacement de ces automates issus directement de l'informatique par un système équivalent fondé sur les circuits des ganglions de la base dédiés à la mémoire à court terme constituerait un raffinement du biomimétisme du modèle.

Enfin, une boucle motrice bien spécifique est dédiée aux mouvements oculomoteurs et cohabite avec un système plus archaïque et plus rapide n'impliquant pas les ganglions de la base et le cortex. Le rôle de cette boucle dans le contrôle du regard, permettant une perception orientée vers l'action future, est à explorer et semble particulièrement important pour la conception de robots dotés de caméras mobiles. La compréhension de la stabilité de ces systèmes constitués de boucles superposées, héritées de l'évolution, par une approche dérivée de la théorie des systèmes dynamiques (Slotine et Lohmiller, 2001) semble être une voie prometteuse.

Conclusion

Du rat de laboratoire...

Nous souhaitons que ce travail, dont les améliorations et perspectives viennent d'être évoquées, puisse constituer une première étape dans la construction d'architectures de contrôle intégrées, inspirées le plus étroitement possible des connaissances neurobiologiques. Comme nous l'avons signalé en introduction, leur implémentation dans un système entier, confronté à un environnement dont les caractéristiques ne lui sont pas données a priori, permet d'envisager à terme des expériences comparatives avec les animaux, afin de mieux évaluer les connaissances biologiques et les modèles qui en découlent.

Nos résultats montrent que les ganglions de la base –et plus particulièrement le noyau accumbens– peuvent être vus, selon l'hypothèse des neurobiologistes, comme un centre dédié à la sélection d'actions intégrant des informations spatiales, motivationnelles et sensorimotrices provenant d'autres structures neurales. En revanche, ils ne répondent pas de manière définitive à la question du codage des informations spatiales en entrée des ganglions de la base, ainsi qu'à celle des niveaux d'interaction entre les différentes boucles des ganglions de la base.

La modélisation d'autres circuits, absents dans la présente architecture et pouvant avoir un impact sur la sélection de l'action –comme ceux permettant une vision active ou une gestion indépendante des motivations– semble une perspective réalisable, compte tenu du bon fonctionnement de l'interaction entre deux boucles réalisé ici.

La possibilité de recueillir les données de n'importe quel élément du modèle ouvre la perspective d'une comparaison avec des données neurophysiologiques concernant l'activation de populations de neurones, les interactions entre différents noyaux, ou diverses modulations synaptiques qui pourront être intégrées dans les futures versions du modèle. Des validations par comparaison d'activation de neurones naturels et artificiels, d'une part, et par comparaison des comportements du robot et du rat de laboratoire, d'autre part, semblent alors raisonnablement envisageables.

...au rat artificiel

Dans le temps imparti à réaliser ce travail, nous n'avons pu que comparer quantitativement notre architecture de contrôle à celles d'animats possédant une sélection de l'action très simplifiée par rapport à l'organisation des ganglions de la base, ou dénués de capacités de navigation. Comme nous l'avons vu, les comparaisons avec d'autres architectures biomimétiques de ce type ne peuvent être que très restreintes actuellement.

Il reste que la confrontation de ces mécanismes, tirés de l'évolution des espèces, avec des mécanismes ingénieurs implémentés dans les nombreux modèles de sélection de l'action que nous avons évoqués en 1.2.2 devrait être menée, dans les mêmes tâches et environnements, afin d'en déterminer plus précisément les avantages et les limitations. L'hypothèse selon laquelle les ganglions de la base réalisent un compromis entre un arbitrage comportemental centralisée et distribué, et combinant des éléments de contrôle à la fois hiérarchiques et non-hiérarchiques dans leur relation avec les autres structures nerveuses (Prescott *et al.*, 1999), pourrait les placer en position d'être comparés à des systèmes ingénieurs de type hiérarchie à libre flux, comme celui de Rosenblatt et Payton (1989), par exemple.

De même, une estimation de l'optimalité des enchaînements comportementaux générés par cette architecture biomimétique pourrait également être envisagée par la confrontation avec des modèles incorporant une approche formelle de la prise de décision, comme par exemple celui de Pirjanian (2000).

Si ces comparaisons, ainsi que d'autres travaux étendant le modèle, démontrent que le rat a pu utilement servir de sources d'inspiration aux informaticiens afin qu'ils dotent leurs robots de plus d'autonomie et d'adaptation, peut-on pour autant prédire l'avènement futur d'un Rongeur Artificiel ? L'intérêt grandissant de plusieurs laboratoires dans la mise en œuvre d'un tel chantier (Cyber rodent (Elfwing *et al.*, 2003), Skinnerbots (Sasksida *et al.*, 1998), AMouse (Lungarella *et al.*, 2002), le rat robot interagissant avec de vrais rats (Takanishi *et al.*, 1998) et Psikharpax (Meyer, 2002)) semble bien augurer d'une telle éventualité.

Annexe A

Abréviations des structures biologiques

- **ACA** : Aire cingulaire antérieure du cortex (rat).
- **AGl** : Cortex agranulaire latéral (rat).
- **AGm** : Cortex agranulaire médian (rat).
- **AIA** : Aire agranulaire insulaire (rat).
- **AVT** : Aire ventrale tégmentale.
- **EP** : Noyau Entopédonculaire (rat).
- **FEF** : Champ frontal oculaire du cortex (primate).
- **FFC** : Force des facteurs causaux.
- **GABA** : Acide gamma-aminobutyrique (neurotransmetteur).
- **GB** : Ganglions de la base.
- **GP** : Globus pallidus, équivalent chez le rat du GPe chez le primate.
- **GPe** : Globus pallidus externe, équivalent chez le primate du GP chez le rat.
- **GPi** : Globus pallidus interne, équivalent chez le primate de l'EP chez le rat.
- **M1** : Aire motrice du cortex (primate).
- **NAcc** : Noyau accumbens.
- **NST** : Noyau subthalamique.
- **MD** : Noyau médio-dorsal du thalamus.
- **MSN** : Neurones épineux moyen.
- **PFC** : Cortex préfrontal.
- **PL/MO** : Cortex prélimbique et médial orbital.
- **PMC** : Cortex prémoteur.
- **S1** : Aire somatosensorielle primaire du cortex.
- **SMA** : Aire motrice supplémentaire du cortex (primate).
- **SNc** : Substance noire compacte.
- **SNr** : Substance noire réticulée.
- **Th** : Thalamus.

- **TRN** : Noyau thalamique réticulé.
- **VA/VL** : Complexe thalamique ventroantérieur-ventrolatéral.
- **VL** : Thalamus ventrolatéral.
- **VM** : Thalamus ventromédian.
- **VP** : Globus pallidus ventral.
- **VPl** : Pallidum ventral, zone dorsale et latérale.
- **VPm** : Pallidum ventral, zone ventrale et médiale.

Annexe B

Tests standards

- **Wisconsin Card Sorting Test (WCST)** : le test de tri de cartes de Wisconsin est une expérience d'apprentissage de catégorisation que l'on mène sur des humains. Le sujet doit trier des cartes qui lui sont présentées. Ces cartes portent toutes de une à trois formes colorées identiques. Il y a trois formes (cercle, rectangle, croix) et trois couleurs (rouge, vert et bleu) possibles. Le tri peut donc être effectué suivant l'un des trois critères suivants : nombre, forme ou couleur des figures. L'expérimentateur, qui attend l'utilisation d'un critère précis, signale verbalement au sujet qu'il fait erreur si le critère retenu par le sujet n'est pas le bon. Le sujet doit alors essayer un autre critère jusqu'à celui qui convient. Le critère retenu par l'expérimentateur peut être modifié plusieurs fois durant l'expérience.
- **Cue Delayed Response Task (cDRT)** : la réponse retardée avec indice de position est une expérience menée sur des primates, composée de deux étapes. Tout d'abord, on présente dans le champ visuel du sujet un objet, à droite ou à gauche, supposé donner un indice. Puis après un court délai, on présente deux objets, l'un à droite, l'autre à gauche, le sujet devant alors pointer celui situé à la position de l'indice. Le sujet est récompensé si la réponse est bonne, ce qui doit lui permettre d'apprendre que le critère important dans l'indice est sa position.
- **Delayed Matching to Sample (DMS)** : cette expérience, semblable dans sa forme à la cDRT, nécessite cependant que le choix final ne soit non pas guidé par la position initiale de l'indice, mais par son aspect : le sujet doit pointer vers l'objet identique à l'indice, quelle que soit sa position.
- **Tâche 2x5** : lors de cette tâche visuo-motrice, un singe est disposé face à seize boutons lumineux disposés en une matrice 4x4. Deux de ces seize boutons sont illuminés simultanément, le singe devant apprendre par essais-erreurs dans quel ordre appuyer sur ces deux boutons. On propose des séquences fixes de cinq de ces tâches. Une séquence est répétée en boucle tant que le singe n'arrive pas à l'exécuter correctement de 10 à 20 fois consécutives, puis l'on passe à une autre séquence. On s'intéresse alors aux temps d'exé-

cution de ces séquences en regard de séquences jamais présentées précédemment, ainsi qu'au nombre d'essais nécessaires à l'acquisition de ces nouvelles séquences.

– **1-2-AX** : La tâche 1-2-AX consiste en la présentation en séquences de symboles alphanumériques (1,2,3,A,B,C,X,Y et Z) devant amener le sujet à presser le bouton droit (R) ou gauche (L). Si le dernier chiffre présenté est le 1, alors toute séquence A-X doit se traduire par l'appui sur R et si le dernier chiffre est 2, c'est la séquence B-Y qui est déclencheuse. Toute séquence autre doit se traduire par le choix de L. On étudie là la capacité du sujet à utiliser sa mémoire à court terme pour se souvenir à la fois du contexte et du début des séquences.
– **Bandit multi-bras** : Dans les problèmes de bandits multi-bras, un agent est mis en présence de *n* bandits-manchots et doit choisir de façon répétée quel bandit-manchot utiliser, sachant que ces machines n'ont pas toutes la même probabilité de rapporter et que ces probabilités ne peuvent être découvertes que par expérimentation.

Annexe C

Paramètres des modèles

C.1 Intégrateurs à fuite

Les neurones artificiels de type intégrateurs à fuite ont une dynamique interne très simple.

Soit X le vecteur d'entrée de dimension n, W celui des poids synaptiques correspondants, l'activation interne du neurone à un instant donné a varie suivant :

$$\frac{d}{dt}a = -k \times (a - \sum_{i=0}^{n} W_i X_i)$$

A partir de cette activation est calculée la sortie y du neurone par passage dans une fonction de transfert linéaire par morceaux, de seuil ϵ et de pente m :

$$y = \begin{cases} 0 & \text{si } a < \epsilon \\ m \times (a - \epsilon) & \text{si } \epsilon \leq a < \epsilon + 1/m \\ 1 & \text{si } \epsilon + 1/m \leq a \end{cases}$$

C.2 Paramètres

Le taux de dopamine influençant le transfert des saliences vers les neurones les sous-parties D1 et D2 du striatum est fixe et vaut $\lambda = 0,2$.

Tous les neurones d'un même module du GPR ont des paramètres identiques. Leurs fonctions de transfert sont linéaires et donc décrites par leurs pentes et seuils, récapitulés dans le tab. C.1.

Les constantes de temps des intégrateurs à fuite du GPR valent toutes $k = 0,25$.

Les interconnexions entre boucles via la voie trans-subthalamique donnent lieu à la définition de deux paramètres, le poids de la connexion du NST de la boucle dorsale vers la SNr de

Module	ϵ	m
Striatum D1	0,2	0,35
Striatum D2	0,2	0,35
NST	-0,25	0,35
GP	-0,2	1
EP/SNr	-0,2	1
Persistance	0	1
TRN	0	0,5
VL	-0,8	0,62

TAB. *C.1: Paramètres des fonctions de transfert des neurones des différents modules du modèle GPR.*

la boucle ventrale, $W_{ib} = 0,4$, et le seuil d'inhibition en sortie de la boucle ventrale en-deçà duquel on considère qu'aucun mouvement n'est généré, $S_{inhib} = -0,3$.

C.3 Calcul des saliences

$Rev(x)$	=	$(1-x)$
$Circ(x)$	=	$\sqrt{1-x^2}$
$f(x)$	=	x^2
$g(x)$	=	$\sqrt{x}$

TAB. *C.2: Fonctions de transfert utilisées en pré-traitemet des calculs de saliences.*

Les calculs de saliences des deux modèles utilisent des fonctions de transfert appliquées aux variables d'entrée. Les fonctions de transfert utilisée sont récapitulées dans le tableau C.2.

C.3.1 Expériences du chapitre 3

Les saliences des expériences de sélection de l'action sont calculées à partir de 4 variables internes :

- L_B : la mesure du degré de *Blancheur* du sol.
- L_N : la mesure du degré de *Noirceur* du sol.
- C_G : la détection d'un *ContactGauche*.
- C_N : la détection d'un *ContactDroit*.

Comportement		Calcul de la salience
ExplorationAléatoire	**WTA**	$-C_G - C_D + 0,5 \times Rev(E_{Pot}) + 0,7 \times Rev(E)$
	GPR	$-C_G - C_D + 0,8 \times Rev(E_{Pot}) + 0,9 \times Rev(E)$
EvitementObstacle	**WTA**	$3C_D + 3C_G$
	GPR	$2C_G + 2C_D + 0.5P_{EO}$
RechargeSurNoir	**WTA**	$-2L_B - C_G - C_D + 3L_N \times Rev(E_{Pot})$
	GPR	$-2L_B - C_G - C_D + 3L_N \times Rev(E_{Pot}) + 0.4P_{RSN}$
RechargeSurBlanc	**WTA**	$-2L_N - C_G - C_D + 5L_B \times Circ(Rev(E_{Pot})) \times Rev(E)$
	GPR	$-2L_N - C_G - C_D + 5L_B \times Circ(Rev(E_{Pot})) \times Rev(E) + 0.5P_{RSB}$
Repos	**WTA**	$-C_G - C_D + 0.1$
	GPR	$-C_G - C_D + 0.6P_R$

TAB. *C.3: Calcul des saliences utilisées par le GPR et le WTA dans les expérimentations de sélection de l'action.*

Complétées de 2 variables internes :

- E : le niveau d'*Energie*.
- E_{Pot} : le niveau d'*Energie Potentielle*.

Et enfin, chaque comportement a accès à la valeur de sa persistence P_{CPT}. Les calculs correspondants figurent dans le tableau C.3.

C.3.2 Expériences du chapitre 4

Les saliences des expériences d'intégration de la navigation et de la sélection de l'action sont calculées à partir de 12 variables externes :

- $\mathbf{Plan}$: le profil de direction combinant les résultats des opérations de planification vers chaque type de ressources, pondérées par les valeurs des motivations correspondantes.
- $\mathbf{RSC}$: le profil de direction indiquant au robot comment retrouver son chemin.
- $\mathbf{Exp}$: le profil de direction indiquant au robot ou explorer.
- $\mathbf{Prox}(\mathbf{res})$: les trois profils de direction indiquant la proximité de chaque type d'objet de l'environnement.
- $maxProx(res)$: les trois maximums des trois profils Prox(res).
- $sur(res)$: trois booléens indiquant si le robot est sur un type d'objet de l'environnement.

Les 4 variables internes sont :

- E : le niveau d'*Energie*.
- E_{Pot} : le niveau d'*Energie Potentielle*.
- Des_L : le degré de désorientation calculé par le système de navigation, lissé.

Boucle ventrale	*Direction i*	$0,65 \times \sqrt{\mathbf{Plan}_i}$
		$+0,55 \times \sqrt{\mathbf{Prox(255)}_i} \times motiv(E)+$
		$+0,55 \times \sqrt{\mathbf{Prox(127)}_i} \times motiv(E_{Pot})$
		$+0,4 \times \mathbf{RSC}_i \times motiv(RSC)$
		$+0,25 \times \mathbf{Exp}_i$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(127)) \times Rev(E_{Pot})$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(255)) \times Rev(E)$
		$+0,2 \times P_{Dir(i)}$
Boucle dorsale	$RechargeE$	$1,2 \times sur(255) \times motiv(E)$
		$+0,6 \times maxProx(255) \times motiv(E) + 0,4P_{RE}$
	$RechargeE_{Pot}$	$1 \times sur(127) \times motiv(E_{Pot})$
		$+0,2 \times maxProx(127) \times motiv(E_{Pot}) + 0,4P_{REp}$

TAB. *C.4: Calculs de saliences utilisés pour les expériences 1, 2 et 3 du chapitre 4. Dans l'expérience 2, des essais seront également réalisés en diminuant le poids associé à la planification (0,65) à 0,55 puis 0,45.*

– $Peur$: le degré d'inhibition des déplacements que génèrent les *Zones Dangereuses*.

Elles sont utilisées pour calculer 4 motivation :

– $motiv(E) = Circ(Rev(E_{Pot})) \times Rev(E)$
– $motiv(E_{Pot}) = Rev(E_{Pot})$
– $motiv(RSC) = Des_L$
– $motiv(ZD) = Peur$

Enfin, chaque canal des boucles ventrales et dorsales a accès à la valeur de sa persistance P_{canal}. Les calculs utilisés dans les diverses expériences sont récapitulés dans les tableaux C.4, C.5 C.6, C.7, C.8 et C.9.

Boucle ventrale	*Direction i*	$+1 \times \sqrt{\mathbf{Prox(255)}_i} \times motiv(E)+$
		$+1 \times \sqrt{\mathbf{Prox(127)}_i} \times motiv(E_{Pot})$
		$+0,35 \times \mathbf{Exp}_i$
		$+0,2 \times P_{Dir(i)}$
Boucle dorsale	$RechargeE$	$1,5 \times sur(255) \times motiv(E) + 0,3P_{RE}$
	$RechargeE_{Pot}$	$1,5 \times sur(127) \times motiv(E_{Pot}) + 0,3P_{REp}$

TAB. *C.5: Calculs de saliences utilisés pour l'animat réactif de l'expérience 1. Il ne tient naturellement pas compte du profil de direction de planification, ni de celui de retour sur ses pas qui ne sert qu'à construire une carte cohérente.*

Boucle ventrale	*Direction i*	$0,45 \times \sqrt{\mathbf{Plan}_i}$
		$+0,35 \times \sqrt{\mathbf{Prox(255)}_i} \times motiv(E)+$
		$+0,35 \times \sqrt{\mathbf{Prox(127)}_i} \times motiv(E_{Pot})$
		$+0,19 \times Rev(\mathbf{Prox(31)}_i) \times motiv(ZD)$
		$+0,4 \times \mathbf{RSC}_i \times motiv(RSC)$
		$+0,05 \times \mathbf{Exp}_i$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(127)) \times Rev(E_{Pot})$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(255)) \times Rev(E)$
		$+0,2 \times P_{Dir(i)}$
Boucle dorsale	$RechargeE$	$1,2 \times sur(255) \times motiv(E)$
		$+0,6 \times maxProx(255) \times motiv(E) + 0,4P_{RE}$
	$RechargeE_{Pot}$	$sur(127) \times motiv(E_{Pot})$
		$+0,2 \times maxProx(127) \times motiv(E_{Pot}) + 0,4P_{REp}$

TAB. *C.6: Calculs de saliences utilisés pour l'expérience 4. Ils intègrent l'utilisation de* $\mathbf{Prox(31)}$ *pour défavoriser les déplacements vers les* Zones Dangereuses.

Boucle ventrale	*Direction i*	$0,55 \times \sqrt{\mathbf{Plan}_i} \times Rev(maxProx(255)) \times Rev(maxProx(127))$
		$+0,55 \times \sqrt{\mathbf{Prox(255)}_i} \times motiv(E)+$
		$+0,55 \times \sqrt{\mathbf{Prox(127)}_i} \times motiv(E_{Pot})$
		$+0,4 \times \mathbf{RSC}_i \times motiv(RSC)$
		$+0,25 \times \mathbf{Exp}_i$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(127)) \times Rev(E_{Pot})$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(255)) \times Rev(E)$
		$+0,2 \times P_{Dir(i)}$
Boucle dorsale	$RechargeE$	$0,9 \times sur(255) \times motiv(E)$
		$+0,1 \times maxProx(255) \times motiv(E) + 0,4P_{RE}$
	$RechargeE_{Pot}$	$0,9 \times sur(127) \times motiv(E_{Pot})$
		$+0,1 \times maxProx(127) \times motiv(E_{Pot}) + 0,4P_{REp}$

TAB. *C.7: Calculs de saliences utilisés pour l'expérience 5. Vis-à-vis de ceux des expériences 1 à 3, le profil de direction issu de la planification se voit modulé par la proximité de ressources visibles d'une part, et les poids de la boucle ventrale ont été diminués d'autre part.*

Boucle ventrale	*Direction i*	$0,15 \times \sqrt{\mathbf{Plan}_i}$
		$0,2 \times \sqrt{\mathbf{Plan}_i} \times Rev(maxProx(255)) \times Rev(maxProx(127))$
		$+0,35 \times \sqrt{\mathbf{Prox(255)}_i} \times motiv(E)+$
		$+0,35 \times \sqrt{\mathbf{Prox(127)}_i} \times motiv(E_{Pot})$
		$+0,13 \times Rev(\mathbf{Prox(31)}_i) \times motiv(ZD)$
		$+0,4 \times \mathbf{RSC}_i \times motiv(RSC)$
		$+0,07 \times \mathbf{Exp}_i$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(127)) \times Rev(E_{Pot})$
		$+0,05 \times \mathbf{Exp}_i \times Rev(maxProx(255)) \times Rev(E)$
		$+0,2 \times P_{Dir(i)}$
Boucle dorsale	$RechargeE$	$0,9 \times sur(255) \times motiv(E)$
		$+0,1 \times maxProx(255) \times motiv(E) + 0,4P_{RE}$
	$RechargeE_{Pot}$	$0,9 \times sur(127) \times motiv(E_{Pot})$
		$+0,1 \times maxProx(127) \times motiv(E_{Pot}) + 0,4P_{REp}$

TAB. *C.8: Calculs de saliences utilisés pour l'expérience 6. L'ensemble des modifications suggérées par les expériences précédentes a été intégré.*

Boucle ventrale	*Direction i*	$+0,5 \times \sqrt{\mathbf{Prox}(\mathbf{255})_i} \times motiv(E)+$
		$+0,5 \times \sqrt{\mathbf{Prox}(\mathbf{127})_i} \times motiv(E_{Pot})$
		$+0,13 \times Rev(\mathbf{Prox}(\mathbf{31})_i) \times motiv(ZD)$
		$+0,25 \times \mathbf{Exp}_i$
		$+0,2 \times P_{Dir(i)}$
Boucle dorsale	$RechargeE$	$0,9 \times sur(255) \times motiv(E)$
		$+0,2 \times maxProx(255) \times motiv(E) + 0,4P_{RE}$
	$RechargeE_{Pot}$	$0,9 \times sur(127) \times motiv(E_{Pot})$
		$+0,2 \times maxProx(127) \times motiv(E_{Pot}) + 0,4P_{REp}$

TAB. *C.9: Calculs de saliences utilisés pour l'animat réactif de l'expérience 6.*

Bibliographie

[Agre et Chapman, 1990] P. Agre et D. Chapman. What are plans for ? *Robotics and autonomous systems*, 6(1-2) :17–24, 1990.

[Alami *et al.*, 1998] R. Alami, R. Chatila, S. Fleury, M. Ghallab, et F. Ingrand. An architecture for autonomy. *International Journal of Robotics Research*, 17(4) :315–377, 1998.

[Albertin *et al.*, 2000] S. Albertin, A. B. Mulder, E. Tabuchi, M. B. Zugaro, et S. I. Wiener. Lesion of the medial shell of the nucleus accumbens impair rats in finding larger rewards, but spare reward-seeking behavior. *Behavioural Brain Research*, 117 :173–183, 2000.

[Albin *et al.*, 1989] R. L. Albin, A. B. Young, et J. B. Penney. The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, 12 :366–375, 1989.

[Albin *et al.*, 1995] R. L. Albin, A. B. Young, et J. B. Penney. The functional anatomy of disorders of the basal ganglia. *Trends in Neurosciences*, 18(2) :63–64, 1995.

[Alexander *et al.*, 1986] G. E. Alexander, M. R. DeLong, et P. L. Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9 :357–381, 1986.

[Alexander *et al.*, 1990] G. E. Alexander, M. D. Crutcher, et M. R. DeLong. Basal ganglia-thalamocortical circuits : Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Progress in Brain Research*, 85 :119–146, 1990.

[Alexander et Crutcher, 1990] G. E. Alexander et M. D. Crutcher. Functional architecture of basal ganglia circuits : neural substrates of parallel processing. *Trends in Neurosciences*, 13 :266–271, 1990.

[Arbib et Dominey, 1995] M. A. Arbib et P. F. Dominey. Modeling the roles of basal ganglia in timing and sequencing saccadic eye movements. In *Models of Information Processing in the Basal Ganglia*, éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser, pages 149–162. The MIT Press, Cambridge, MA, 1995.

[Arbib et Lieblich, 1977] M. A. Arbib et I. Lieblich. Motivational learning of spatial behavior. In *Systems Neuroscience*, éditeur J. Metzler, pages 221–239. Academic Press, New York, 1977.

[Arbib, 1995] M. Arbib. Introducing the neuron. In *The handbook of brain theory and neural networks*, éditeur M. Arbib. The MIT Press, Cambridge, MA, 1995.

[Arleo et Gerstner, 2000] A. Arleo et W. Gerstner. Spatial cognition and neuro-mimetic navigation : a model of hippocampal place cell activity. *Biological Cybernetics*, 83 :287–299, 2000.

[Arleo, 2000] A. Arleo. *Spatial learning and navigation in neuro-mimetic systems, modelling the rat hippocampus*. PhD thesis, Swiss Federal Institute of technology, EPFL, Switzerland, 2000.

[Ashby, 1952] W. R. Ashby. *Design for a brain*. Chapman and Hall, 1952.

[Atkinson et Birch, 1970] J. W. Atkinson et D. Birch. *The dynamics of action*. John Wiley & Sons, 1970.

[Baddeley, 1993] A. Baddeley. *La mémoire humaine : théorie et pratique*. Presse Universitaire de Grenoble, 1993.

[Baldassare, 2003] G. Baldassare. Computational behavioneuroscience. Rapport technique, Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Rome, 2003.

[Baldassarre, 2001] G. Baldassarre. A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviours. *Cognitive Sytems Research*, 2001.

[Balkenius et Moren, 2000] C. Balkenius et L. Moren. A computational model of context processing. In *From animals to animats 6 : Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, éditeurs J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, et S. Wilson, pages 256–265, Cambridge, MA, 2000. The MIT Press.

[Bar-Cohen et Breazeal, 2003] éditeurs Y. Bar-Cohen et C. Breazeal. *Biologically inspired intelligent robots*, volume PM 122. SPIE Press, 2003.

[Bar-Gad *et al.*, 2000] I. Bar-Gad, G. Havazelet Heimer, J. A. Goldberg, E. Ruppin, et H. Bergman. Reinforcement driven dimensionality reduction – a model for information processing in the basal ganglia. *Journal of Basic and Clinical Physiology and Pharmacology*, 11 :305–320, 2000.

[Bar-Gad et Bergman, 2001] I. Bar-Gad et H. Bergman. Stepping out of the box : information processing in the neural networks of the basal ganglia. *Current Opinion in Neurobiology*, 11 :689–695, 2001.

[Barto, 1995] A. G. Barto. Adaptive critics and the basal ganglia. In *Models of Information Processing in the Basal Ganglia*, éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser, pages 215–232. The MIT Press, Cambridge, MA, 1995.

[Beer et Chiel, 1991] R. D. Beer et H. J. Chiel. The neural basis of behavioral choice in an artificial insect. In *From animals to animats : proceedings of the conference of adaptive behavior*, éditeurs J.-A. Meyer et S. W. Wilson, pages 247–254, Cambridge, MA, 1991. MIT Press.

[Beiser *et al.*, 1997] D. G. Beiser, S. E. Hua, et J. C. Houk. Network models of the basal ganglia. *Current Opinion in Neurobiology*, 7 :185–190, 1997.

[Beiser et Houk, 1998] D. G. Beiser et J. C. Houk. Model of cortical-basal ganglionic processing : encoding the serial order of sensory events. *Journal of Neurophysiology*, 79 :3168–3188, 1998.

[Berns et Sejnowski, 1996] G. S. Berns et T. J. Sejnowski. How the basal ganglia make decision. In *The neurobiology of decision making*, éditeurs A. Damasio, H. Damasio, et Y. Christen, pages 101–113. Springer-Verlag, Berlin, 1996.

[Berns et Sejnowski, 1998] G. S. Berns et T. J. Sejnowski. A computational model of how the basal ganglia produce sequences. *Journal of Cognitive Neuroscience*, 10(1) :108–121, 1998.

[Berridge et Robinson, 1998] K. C. Berridge et T. E. Robinson. What is the role of dopamine in reward : hedonic impact, reward learning , or incentive salience. *Brain Research Reviews*, 28 :309–369, 1998.

[Berthoz, 2003] A. Berthoz. *La Décision*. Odile jacob, 2003.

[Bieszczad, 1994] A. Bieszczad. Neurosolver : A step toward a neuromorphic general problem solver. In *Proceedings of IEEE World Congress on Computational Intelligence*, volume III, pages 1313–1318, Orlando, FL, 1994. IEEE Press.

[Bischoff, 1998] A. Bischoff. *Modelling the basal ganglia in the control of arm movements*. PhD thesis, University of Southern California, 1998.

[Blumberg, 1994] B. Blumberg. Action-selection in hamsterdam : Lessons from ethology. In *From Animals To Animats 3 : Proceedings of the Third International Conference on the Simulation of Adaptive Behavior*, éditeurs D. Cliff, P. Husbands, J.-A. Meyer, et S.W. Wilson, Cambridge, MA, 1994. MIT Press.

[Blumberg, 1996] B. Blumberg. *Old tricks, new dogs : ethology and interactive creatures*. PhD thesis, MIT Media Lab, 1996.

[Bongard et Paul, 2000] J. C. Bongard et C. Paul. Investigationg morphological symmetry and locomotive efficiency using virtual embodied evolution. In *From animals to animats 6 : proceedings of the sixth conference of adaptive behavior*, éditeurs J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, et S. Wilson, page à retrouver, Cambridge, MA, 2000. MIT Press.

[Brooks, 1986] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1) :14–23, 1986.

[Brooks, 1991] R. A. Brooks. Intellligence without reason. In *Proceedings of 12th International Joint Conference on Artifi cial Intelligence*, éditeurs J. Myopoulos et R. Reiter, pages 569–595, San Mateo, CA, 1991. Morgan Kaufmann publishers Inc.

[Brown *et al.*, 1999] J. Brown, D. Bullock, et S. Grossberg. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. Rapport Technique CAS/CNS-TR-99-011, Boston University, Department of Cognitive and Neural Systems and Center for Adaptive Systems, 1999.

[Bryson, 2000] J. Bryson. Hierarchy and sequence vs. full parallelism in action selection. In *From animals to animats 6 : Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, éditeurs J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, et S. W. Wilson, volume 1, pages 147–156, Cambridge, MA, 2000. The MIT Press.

[Burgess *et al.*, 1997] N. Burgess, J. G. Donnett, K. J. Jeffrey, et J. O'Keefe. Robotic and neuronal simulation of the hippocampus and rat navigation. *Philosophical Transactions of the Royal Society*, 352 :1535–1543, 1997.

[Burnod, 1989] Y. Burnod. *An adaptive neural network : the cerebral cortex*. Masson, 1989.

[Calhoun, 1962] J. B. Calhoun. The ecology and sociology of the norway rat. Rapport Technique 1008, U S Dept of Health, Education and Welfare Public Health Service, 1962.

[Cardinal, 2001] R. N. Cardinal. *Neuropsychology of reinforcement processes in the rat*. PhD thesis, University of Cambridge, 2001.

[Chatila, 2002] R. Chatila. Evaluer l'autonomie, mais comment ? Rapport Technique 02092, LAAS, France, 2002.

[Chevalier et Deniau, 1990] G. Chevalier et M. Deniau. Disinhibition as a basic process of striatal functions. *Trends in Neurosciences*, 13 :277–280, 1990.

[Connolly et Burns, 1993] C. I. Connolly et J. B. Burns. A model for the functionning of the striatum. *Biological Cybernetics*, 68 :535–544, 1993.

[Connolly et Burns, 1995] C. I. Connolly et J. B. Burns. A state-space striatal model. In *Models of Information Processing in the Basal Ganglia*, éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser, pages 163–177. The MIT Press, Cambridge, MA, 1995.

[Contreras-Vidal et Schultz, 1999] J. L. Contreras-Vidal et W. Schultz. A predictive reinforcement model of dopamine neurons for learning approach behavior. *Journal of Computational Neuroscience*, 6(3) :191–214, 1999.

[Crabbe, 2002] F. L. Crabbe. Compromise candidates in positive goal scenarios. In *From animals to animats 7 : Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, éditeurs B. Hallam, D. Floreano, J. Hallam, G. Hayes, et J.-A. Meyer, pages 105–106, Cambridge, MA, 2002. MIT Press.

[Daw *et al.*, 2002] N. D. Daw, D. S. Touretzky, et W. E. Skaggs. Effects of reward type and changing task demands on striatal representation in the rat. *Society of Neuroscience Abstracts*, 28 :765, 2002.

[Dawkins, 1995] R. Dawkins. *River our of Eden : A darwinian view of life*. Basic books, 1995.

[Dayan, 2001] P. Dayan. Motivated reinforcement learning. In *Neural Information Processing Systems*, éditeurs Todd Leen, Tom Dietterich, et Volker Tresp, volume 13. The MIT Press, Cambridge, MA, 2001.

[Deniau, 2003] J. M. Deniau. Communication personnelle, 2003.

[Dominey et Arbib, 1992] P. F. Dominey et M. A. Arbib. A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex*, 2 :153–175, 1992.

[Doya, 1999] K. Doya. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex ? *Neural Networks*, 12 :961–974, 1999.

[Doya, 2000] K. Doya. Complementary roles of the basal ganglia and the cerebellum in learning and motor control. *Current opinion in neurobiology*, 10(6) :732–739, 2000.

[Dreyfus, 1972] H. Dreyfus. *What computers can't do : a critique of artifi cial reason*. MIT Press, Cambridge, MA, 1972.

[Elfwing *et al.*, 2003] S. Elfwing, E. Uchibe, et K. Doya. An evolutionary approach to automatic construction of the structure in hierarchical reinforcement learning. In *Genetic and evolutionary computation conference (GECCO)*, pages 507–509, 2003.

[Everitt et Wolf, 2002] B. J. Everitt et M. E. Wolf. Psychomotor stimulant addiction : a neural systems perspective. *Journal of neuroscience*, 22(9) :3312–3320, 2002.

[Filliat, 2001] D. Filliat. *Cartographie et estimation globale de la position pour un robot mobile autonome*. PhD thesis, LIP6/AnimatLab, Université Paris 6, France, 2001.

[Floresco *et al.*, 1997] S. B. Floresco, J. K. Seamans, et A. G. Phillips. Selective roles for hippocampal, prefrontal cortical, and ventral striatal circuits in radial-arm maze tasks with or without a delay. *Journal of neuroscience*, 17 :1880–1890, 1997.

[Florian, 2003] R. V. Florian. Autonomous artificial intelligent agents. Rapport Technique Coneural-03-01, Center for cognitive and neural studies, Romania, 2003.

[Franceschini *et al.*, 1992] N. Franceschini, J. M. Pichon, et C. Blanes. From insect vision to robot vision. *Philosophical Transactions of the Royal Society of London*, 337 :283–294, 1992.

[Frank *et al.*, 2000] M. J. Frank, B. Loughry, et R. C. O'Reilly. Interactions between frontal cortex and basal ganglia in working memory : a computational model. *Cognitive, Affective and Behavioral Neuroscience*, 1 :137–160, 2000.

[Frezza-Buet et Alexandre, 1999] H. Frezza-Buet et F. Alexandre. Modeling prefrontal functions for robot navigation. In *International joint conference on neural networks*, 1999.

[Fuster, 1989] J. Fuster. *The prefrontal cortex*. Raven Press, New York, 1989.

[Gat, 1991] E. Gat. *Reliables Goal-Directed Reactive Control of Autonomous Mobile Robot*. PhD thesis, Virginia Polytechnic Institue and State University, 1991.

[Gaussier *et al.*, 2000] P. Gaussier, S. Leprêtre, M. Quoy, A. Revel, C. Joulain, et J.-P. Banquet. Experiments and models about cognitive map learning for motivated navigation. In *Interdisciplinary approaches to robot learning*, éditeur J. Demiris & A. Birk, volume 24 de *Robotics and Intelligent Systems*, pages 53–94. World Scientific, 2000.

[Gelfand *et al.*, 1997] J. Gelfand, V. Gullapalli, M. Johnson, C. Raye, et J. Henderson. The dynamics of prefrontal cortico-thalamo-basal ganglionic loops and short-trm memory interference phenomena. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, pages 253–258, 1997.

[Gibson, 1966] J. J. Gibson. *The Senses Considered as Perceptual Systems*. Allen and Unwin, 1966.

[Gillies et Arbruthnott, 2000] A. Gillies et G. Arbruthnott. Computational models of the basal ganglia. *Movement Disorders*, 15(5) :762–770, 2000.

[Girard *et al.*, 2001] B. Girard, G. Robert, et A. Guillot. Jeu vidéo et intelligence artificielle située. *In Cognito*, 22, 2001.

[Girard *et al.*, 2002] B. Girard, V. Cuzin, A. Guillot, K. N. Gurney, et T. J. Prescott. Comparing a bio-inspired robot action selection mechanism with winner-takes-all. In *From Animals to Animats 7 : Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, éditeurs B. Hallam, D. Floreano, J. Hallam, G. Hayes, et J.-A. Meyer. The MIT Press, 2002.

[Girard *et al.*, 2003a] B. Girard, V. Cuzin, A. Guillot, K. N. Gurney, et T. J. Prescott. A basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of integrative neuroscience*, 2003. Accepté sous réserve de modifications.

[Girard *et al.*, 2003b] B. Girard, D. Filliat, J.-A. Meyer, A. Berthoz, et A. Guillot. An integration of two control architectures of action selection and navigation inspired by neural circuits in the vertebrate : the basal ganglia. In *Eighth Neural Computation and Psychology Workshop (NCPW 8) Connectionist Models of Cognition, Perception and Emotion*, Progress in Neural Processing. World Scientific, 2003. Accepté sous réserve de modifications.

[Goldman-Rakic, 1994] P. Goldman-Rakic. The circuitry of working memory revealed by anatomy and metabolic imaging. In *Motor and cognitive functions of the prefrontal cortex*, éditeurs A.M. Thierry, J. Glowinski, P. Goldman-Rakic, et Y. Christen. Springer-Verlag, Berlin, 1994.

[Gourichon *et al.*, 2002] S. Gourichon, J.-A. Meyer, et P. Pirim. Using coloured snapshots for short-range guidance in mobile robots. *International Journal of Robotics and Automation*, 17(4) :154–162, 2002.

[Graybiel, 1998] A. M. Graybiel. The basal ganglia and chunking of action repertories. *Neurobiology of Learning and Memory*, 70 :119–136, 1998.

[Greenberg, 2001] N. Greenberg. The past and future of the basal ganglia. In *The neuroethology of Paul MacLean : frontiers and convergences*, éditeurs G. Cory et R. Gardner. Praeger, 2001.

[Groenewegen *et al.*, 1996] H. J. Groenewegen, C. I. Wright, et A. V. J. Beijer. The nucleus accumbens : gateway for limbic stuctures to reach the motor system ? *Progress in Brain Research*, 107 :485–511, 1996.

[Groenewegen *et al.*, 1999] H. J. Groenewegen, A. B. Mulder, A. V. J. Beijer, C. I. Wright, F. H. Lopes Da Silva, et C. M. A. Pennartz. Hippocampal and amygdaloid interactions in the nucleus accumbens. *Psychobiology*, 27(2) :149–164, 1999.

[Guazzelli *et al.*, 1998] A. Guazzelli, F. J. Corbacho, M. Bota, et M. A. Arbib. Affordances, motivations and the worls graph theory. *Adaptive Behaviour : special issue on biologically inspired models of spatial navigation*, 6(3/4) :435–471, 1998.

[Guillot et Meyer, 2000] A. Guillot et J.-A. Meyer. From sab94 to sab2000 : What's new, animat ? In *From animals to animats 6 : Proceedings of the sixth international conference on simulation of adaptive behavior*, éditeurs J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, et S. W. Wilson, pages 3–12, 2000.

[Guillot et Meyer, 2001] A. Guillot et J.-A. Meyer. The animat contribution to cognitive systems research. *Journal of Cognitive Systems Research*, 2(2) :157–165, 2001.

[Guillot et Meyer, 2002] A. Guillot et J.-A. Meyer. Psikharpax, l'ambition d'être un rat. *La Recherche. Numéro spécial : les nouveaux robots*, 350 :64–67, 2002.

[Guillot et Meyer, 2003] A. Guillot et J.-A. Meyer. La contribution de l'approche animat aux sciences cognitives. *Cahiers romans de sciences cognitives*, 1(1) :1–26, 2003.

[Guillot, 1986] A. Guillot. Revue générale des méthodes d'étude des séquences comportementales. In *Etudes et analyses comportementales*, éditeur A. Guillot, volume 2, pages 86–106. Groupe d'Etude du Comportement, 1986.

[Guillot, 1988] A. Guillot. *Contribution à l'étude des séquences comportementales de la souris : approches descriptive, causale et fonctionnelle*. PhD thesis, University of Paris 7, 1988.

[Gurney *et al.*, 2001a] K. Gurney, T. J. Prescott, et P. Redgrave. A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological Cybernetics*, 84 :401–410, 2001.

[Gurney *et al.*, 2001b] K. Gurney, T. J. Prescott, et P. Redgrave. A computational model of action selection in the basal ganglia. ii. analysis and simulation of behaviour. *Biological Cybernetics*, 84 :411–423, 2001.

[Gurney, 1992] K. Gurney. Training nets of hardware realizable sigma-pi units. *Neural Networks*, 5(2) :289–303, 1992.

[Heimer *et al.*, 1997] L. Heimer, G.F. Alheid, J.S. de Olmos, H. Groenewegen, S. Haber, R.E. Harlan, et D. Zahm. The accumbens : beyon the core-shell dichotomy. *Journal of neuropsychiatry*, 9 :354–381, 1997.

[Hikosaka *et al.*, 1999] O. Hikosaka, H. Nakahara, M. K. Rand, K. Sakai, X. Lu, K. Nakamura, S. Miyachi, et K. Doya. Parallel neural networks for learning sequential procedures. *Trends in Neurosciences*, 22(10) :464–471, 1999.

[Holscher *et al.*, 2003] C. Holscher, W. Jacob, et H. A. Mallot. Reward modulates neuronal activity in the hippocampus of the rat. *Behavioral brain research*, 142(1-2) :181–191, 2003.

[Honkanen, 1999] A. Honkanen. *Modulation of Brain Dopaminergic neurotransmission in alcohol-preferring rats by alcohol and opioids*. PhD thesis, Division of Pharmacology and Toxicology, Department of Pharmacy, University of Helsinki, 1999.

[Houk *et al.*, 1995a] J. C. Houk, J. L. Adams, et A. G. Barto. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In *Models of Information Processing in the Basal Ganglia*, éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser, pages 249–271. The MIT Press, Cambridge, MA, 1995.

[Houk *et al.*, 1995b] éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser. *Models of Information Processing in the Basal Ganglia*. The MIT Press, Cambridge, MA, 1995.

[Houk et Wise, 1995] J. C. Houk et S. P. Wise. Distributed modular architectures linking basal ganglia, cerebellum and cerebral cortex : their role in planning and controlling action. *Cerebral Cortex*, 5 :95–110, 1995.

[Houston et Sumida, 1985] A. Houston et B. Sumida. A positive feedback model for switching between two activities. *Animal Behaviour*, 33 :315–325, 1985.

[Humphries et Gurney, 2002] M. D. Humphries et K. N. Gurney. The role of intra-thalamic and thalamocortical circuits in action selection. *Network : Computation in Neural Systems*, 13 :131–156, 2002.

[Humphries, 2002] M. D. Humphries. *The basal ganglia and action selection : A computational study at multiple levels of description*. PhD thesis, University of Sheffield, 2002.

[Ijspeert, 2001] A. J. Ijspeert. A connectionist central pattern generator for the aquatic and terrestrial gaits of a simulated salamander. *Biological cybernetics*, 84(5) :331–348, 2001.

[Ikemoto et Panksepp, 1999] S. Ikemoto et J. Panksepp. The role of nucleus accumbens dopamine in motivated behavior : a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, 31 :6–41, 1999.

[Ikemoto, 2002] S. Ikemoto. Ventral striatal anatomy of locomotor activity induced by cocaine, d-amphetamine, dopamine and d_1/d_2 agonists. *Neuroscience*, 113(4) :939–955, 2002.

[Immelmann, 1980] K. Immelmann. *Introduction to ethology*. Plenum Press, 1980.

[Jackson et Houghton, 1992] S. Jackson et G. Houghton. Basal ganglia function in the control of visuospatial attention : A neural-network model. Rapport Technique 92-6, University of Oregon, Institute of Cognitive and Decision Sciences, 1992.

[Jackson et Houghton, 1995] S. Jackson et G. Houghton. Sensorimotor selection and the basal ganglia : A neural network model. In *Models of Information Processing in the Basal Ganglia*, éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser, pages 337–367. The MIT Press, Cambridge, MA, 1995.

[Jaeger *et al.*, 1994] D. Jaeger, H. Kita, et C. J. Wilson. Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *Journal of Neurophysiology*, 72 :2555–2558, 1994.

[Joel *et al.*, 2002] D. Joel, Y. Niv, et E. Ruppin. Actor-critic models of the basal ganglia : new anatomical and computational perspectives. *Neural Networks*, 15(4–6), 2002.

[Joel et Weiner, 1994] D. Joel et I. Weiner. The organization of the basal ganglia-thalamocortical circuits : open interconnected rather than closed segregated. *Neuroscience*, 63 :363–379, 1994.

[Joel et Weiner, 2000] D. Joel et I. Weiner. The connections of the dopaminergic system with the striatum in rats and primates : An analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, 96(3) :452–474, 2000.

[Kaelbling *et al.*, 1996] L. P. Kaelbling, M. L. Littman, et A. W. Moore. Reinforcement learning : a survey. *Journal of Artifi cial Intelligence Research*, 4 :237–285, 1996.

[Keijzer, 1998] F. A. Keijzer. Some armchair worries about wheeled behavior. In *From Animals to Animats 5 : Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, éditeurs R. Pfeifer, B. Blumberg, J-A. Meyer, et S.W. Wilson, pages 13–21, Cambridge, MA, 1998. MIT Press.

[Kelley, 1999] A. E. Kelley. Neural integrative activities of nucleus accumbens subregions in relation to learning and motivation. *Psychobiology*, 27 :198–213, 1999.

[Khamassi *et al.*, 2003] M. Khamassi, B. Girard, A. Guillot, et A. Berthoz. Mécanismes neuromimétiques d'apprentissage par renforcement dans l'architecture de contrôle du rat artificiel psikharpax. Poster présenté à la plate-forme AFIA, 2003.

[Khamassi, 2003] M. Khamassi. Un modèle d'apprentissage par renforcement dans une architecture de contrôle de la sélection de l'action chez le rat artificiel psikharpax. Master's thesis, Université Paris 6, France, 2003.

[Kobayashi *et al.*, 2003] T. Kobayashi, A. H. Tran, H. Nishijo, T. Ono, et G. Matsumoto. Contribution of hippocampal place cell activity to learning and formation of goal-directed navigation in rats. *Neuroscience*, 117(4) :1025–1035, 2003.

[Kolomiets *et al.*, 2001] B. P. Kolomiets, J.-M. Deniau, P. Mailly, A. Menetreyand J. Glowinski, et A.-M. Thierry. Segregation and convergence of information flow through the cortico-subthalamic pathways. *Journal of Neuroscience*, 21(15) :5764–5772, 2001.

[Kolomiets *et al.*, 2003] B. P. Kolomiets, J. M. Deniau, J. Glowinski, et A.-M. Thierry. Basal ganglia and processing of cortical information : functional interactions between trans-striatal and trans-subthalamic circuits in the substantia nigra pars reticulata. *Neuroscience*, 117(4) :931–938, 2003.

[Korenzy et Borenstein, 1991] Y. Korenzy et J. Borenstein. Potential fields methods and their inherent limitations for mobile robot navigation. In *Proceedings of the IEEE internatinal conference on robotics and automation*, pages 489–493, 1991.

[Laithier *et al.*, 2002] S. Laithier, B. Girard, V. Cuzin, A. Guillot, et T. Prescott. A vertebrate brain-inspired model of action selection : More than a winner-takes-all ? Poster presented at Robotics as Theoretical Biology Workshop, 2002.

[Lambrinos *et al.*, 2000] D. Lambrinos, R. Moller, T. Labhart, R. Pfeiffer, et R. Wehner. A mobile robot employing insect strategies for navigation. *Robotics and Autonomous Systems*, 30 :39–64, 2000.

[Lorenz, 1950] K. Z. Lorenz. The comparative method in studying innate behaviour patterns. In *Symposia of the Society for Experimental Biology*, numéro 4, pages 221–268, Cambridge, 1950. Cambridge University Press.

[Lungarella *et al.*, 2002] M. Lungarella, V. V. Hafner, R. Pfeifer, et H. Yokoi. Whisking : An unexplored sensory modality. In *From animals to animats 7 : Proceedings of the Seventh International Conference on the Simulation of Adaptive Behavior*, pages 58–59, Cambridge, MA, 2002. MIT Press.

[Lörincz, 1997] A. Lörincz. Neurocontrol iii : differencing models of basal ganglia thalamo-cortical loops. *Neural Network World*, 7 :43–72, 1997.

[Maes, 1989] P. Maes. How to do the right thing. *Connection Science Journal*, 1(3) :291–323, 1989.

[Maes, 1991] P. Maes. A bottom-up architecture for behavior selection in an artificial creature. In *From animals to animats : Proceedings of the First International Conference on Simula-*

tion of Adaptive Behavior, éditeurs J.-A. Meyer et S. Wilson, pages 478–485, Cambridge, MA, 1991. MIT Press.

[Martin et Ono, 2000] P. D. Martin et T. Ono. Effects of reward anticipation, reward presentation, and spatial parameters on the firing of single neurons recorded in the subiculum and nucleus accumbens of freely moving rats. *Behavioural Brain research*, 116(1) :23–28, 2000.

[Maurice *et al.*, 1997] N. Maurice, J.-M. Deniau, A. Menetrey, J. Glowinski, et A.-M. Thierry. Position of the ventral pallidum in the rat prefrontal cortex-basal ganglia circuit. *Neuroscience*, 80(2) :523–534, 1997.

[Maurice *et al.*, 1999] N. Maurice, J.-M. Deniau, J. Glowinski, et A.-M. Thierry. Relationships between the prefrontal cortex and the basal ganglia in the rat : physiology of the cortico-nigral circuits. *Journal of neuroscience*, 19(11) :4674–4681, 1999.

[McFarland et Sibly, 1975] D. J. McFarland et R. M. Sibly. The behavioural final common path. *Philosophical Transactions of the Royal Society of London*, 270 :265–293, 1975.

[McFarland, 1971a] D. McFarland. *Feedback mechanisms in animal behaviour*. Academic Press, London, 1971.

[McFarland, 1971b] D. J. McFarland. *Feedback mechanisms in animal behaviour*. Academic Press, London, 1971.

[McFarland, 1977] D. J. McFarland. Decision making in animals. *Nature*, 269 :15–21, 1977.

[McFarland, 1989] D. McFarland. *Animal behaviour problems*. Addison-Wesley Longman, ltd., 1989.

[McGeorge et Faull, 1989] A. J. McGeorge et R. L. Faull. The organization of the projections from the cerebral cortex to the striatum in the rat. *Neuroscience*, 29 :503–537, 1989.

[McNaughton *et al.*, 1993] B. L. McNaughton, C. A. Barnes, J. L. Gerrard, K. Gothard, M. W. Jung, J. J. Knierim, H. Kudrimoti, Y. Qin, W. E. Skaggs, M. Suster, et K. L. Weaver. Deciphering the hippocampal polyglot : the hippocampus as a path integration system. *Journal of Experimental Biology*, 199 :173–185, 1993.

[Mel, 1993] B. W. Mel. Synaptic integration in an excitable dendritic tree. *Journal of Neurophysiology*, 70(3) :1086–1101, 1993.

[Meyer et Guillot, 1991] J.-A. Meyer et A. Guillot. Simulation of adaptive behavior in animats : review and prospect. In *From animals to animats : Proceedings of the First International Conference on the Simulation of Adaptive Behavior*, éditeurs J. A. Meyer et S. W. Wilson, Cambridge, MA, 1991. The MIT Press/Bradford Books.

[Meyer et Guillot, 1994] J.-A. Meyer et A. Guillot. From sab90 to sab94 : Four years of animat research. In *From animals to animats 3 : Proceedings of the third international conference*

on simulation of adaptive behavior, éditeurs D. Cliff, P. Husbands, J.-A. Meyer, et S. W. Wilson. The MIT Press/Bradford Books, 1994.

[Meyer, 1996] J.-A. Meyer. Pour une approche complémentaire de l'ia traditionnelle : le manifeste animat. *In Cognito*, 6 :1–4, 1996.

[Meyer, 2002] J.-A. Meyer. Vers la synthèse d'un rat artificiel. In *Actes des journées du programme interdisciplinaire ROBEA*, pages 29–32. Publications LAAS, 2002.

[Middleton et Strick, 1994] F. A. Middleton et P. L. Strick. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, 266 :458–461, 1994.

[Mink, 1996] J. W. Mink. The basal ganglia : Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50(4) :381–425, 1996.

[Monchi *et al.*, 2000] O. Monchi, J. G. Taylor, et A. Dagher. A neural model of working memory processesin normal subjects, parkinson's disease and schizophrenia for fmri design and predictions. *Neural Networks*, 13 :953–973, 2000.

[Montague *et al.*, 1996] P. R. Montague, P. Dayan, et T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, 16(5) :1936–1947, 1996.

[Montes-Gonzalez *et al.*, 2000] F. Montes-Gonzalez, T. J. Prescott, K. N. Gurney, M. Humphries, et P. Redgrave. An embodied model of action selection mechanisms in the vertebrate brain. In *From animals to animats 6 : Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, éditeurs J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, et S. W. Wilson, volume 1, pages 157–166, Cambridge, MA, 2000. The MIT Press.

[Montes-Gonzalez, 2001] F. Montes-Gonzalez. *A robot model of action selection in the vertebrate brain*. PhD thesis, University of Sheffield, UK, 2001.

[Morgan, 1894] C. L. Morgan. *An introduction to comparative psychology*. E. Arnold, London, 1894.

[Mulder *et al.*, submitted] A. B. Mulder, E. Tabuchi, et S. I. Wiener. Striatal neurons parse displacement sequences to engage hippocampal maps for navigation. *Nature neuroscience*, submitted.

[Nakahara *et al.*, 2001] H. Nakahara, K. Doya, et O. Hikosaka. Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences : A computational aproach. *Journal of Cognitive Neuroscience*, 13 :626–647, 2001.

[Parent et Hazrati, 1995a] A. Parent et L.-N. Hazrati. Functional anatomy of the basal ganglia. i. the cortico-basla ganglia-thalamo-cortical loop. *Brain Research Reviews*, 20 :91–127, 1995.

[Parent et Hazrati, 1995b] A. Parent et L.-N. Hazrati. Functional anatomy of the basal ganglia. ii. the place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Research Reviews*, 20 :128–154, 1995.

[Pennartz *et al.*, 1994] C. M. A. Pennartz, H. J. Groenewegen, et F. H. Lopes Da Silva. The nucleus accumbens as a complex of functionally distinct neuronal ensembles : an integration of behavioural, electrophysiological and anatomical data. *Progress in neurobiology*, 42 :719–761, 1994.

[Pennartz *et al.*, 2000] C. M. A. Pennartz, B. L. McNaughton, et A. B. Mulder. *Progress in Brain Research*, volume 126, chapitre The glutamate hypothesis of reinforcement, pages 231–253. Elsevier Science, 2000.

[Pennartz, 1996] C. M. A. Pennartz. The ascending neuromodulatory systems in learning by reinforcement : comparaing computational conjectures with experimental findings. *Brain Research Reviews*, 21 :219–245, 1996.

[Pennartz, 1997] C. M. A. Pennartz. Reinforcement learning by hebbian synapses with adaptive threshold. *Neuroscience*, 81 :303–319, 1997.

[Pirjanian, 1997] P. Pirjanian. An overview of system architectures for action selection in mobile robots. Rapport technique, Laboratory of Image Analysis, Aalborg University, 1997.

[Pirjanian, 1999] P. Pirjanian. Behavior coordination mechanisms – state-of-the-art. Rapport Technique IRIS-99-375, Institute of Robotics and Intelligent Systems, School of Engineering, University of Southern California, 1999.

[Pirjanian, 2000] P. Pirjanian. Multiple objective behavior-based control. *Robotics and autonomous systems*, 31(1-2), 2000.

[Prescott *et al.*, 1999] T. J. Prescott, P. Redgrave, et K. N. Gurney. Layered control architectures in robot and vertebrates. *Adaptive Behavior*, 7(1) :99–127, 1999.

[Prescott, 2001] T. J. Prescott. The evolution of action selection. In *The whole iguana*, éditeurs O. Holland et D. McFarland. MIT Press, Cambridge, MA, 2001.

[Preuss, 1995] T. M. Preuss. Do rats have prefrontal cortex ? *Journal of cognitive neuroscience*, 7 :1–24, 1995.

[Quoy *et al.*, 2002] M. Quoy, P. Laroque, et P. Gaussier. Learning and motivational couplings promote smarter behaviors of an animat in an unknown world. *Robotics and autonomous systems*, 38 :149–156, 2002.

[Redgrave *et al.*, 1999a] P. Redgrave, T. J. Prescott, et K. Gurney. The basal ganglia : a vertebrate solution to the selection problem ? *Neuroscience*, 89(4) :1009–1023, 1999.

[Redgrave *et al.*, 1999b] P. Redgrave, T. J. Prescott, et K. N. Gurney. Is the short-latency dopamine response too short to signal reward error ? *Trends in neuroscience*, 22 :146–151, 1999.

[Rolls, 1999] E. T. Rolls. *The brain and emotion*. Department of Experimental Psychology, Oxford University, 1999.

[Romo et Schultz, 1990] R. Romo et W. Schultz. Dopamine neurons of the monkey midbrain : contingencies of response to active touch during self-initiated movements. *Journal of Neurophysiology*, 63 :592–606, 1990.

[Roper et Crossland, 1982] T. J. Roper et G. Crossland. Mechanisms underlying eating-drinking transitions in rats. *Animal Behaviour*, 30 :602–614, 1982.

[Rosenblatt et Payton, 1989] J. K. Rosenblatt et D. Payton. A fine-grained alternative to the subsumption architecture for mobile robot control. In *Proceedings of the IEEE/INNS international joint conference on neural networks*, 1989.

[Rosenblatt, 1995] J. K. Rosenblatt. Damn : A distributed architecture for mobile navigation. In *AAAI spring symposium on lessons learned from implemented software architectures for physical agents*, Menlo Park, CA, 1995. AAAI Press.

[Rumelhart et McClelland, 1986] D. Rumelhart et J. McClelland. *Parallel distributed processing*. The MIT Press, Cambridge, MA, 1986.

[Salum *et al.*, 1999] C. Salum, A. Roque da Silva, et A. Pickering. Striatal dopamine in attentional learning : a computational model. *Neurocomputing*, 26-27 :845–854, 1999.

[Sasksida *et al.*, 1998] L. M. Sasksida, S. M. Raymond, et D. S. Touretsky. Shaping robot behavior using principles from instrumental conditioning. *Robotics and Autonomous Systems*, 22(3-4) :231–249, 1998.

[Schultz *et al.*, 1997] W. Schultz, P. Dayan, et P. R. Montague. A neural substrate of prediction and reward. *Science*, 275 :1593–1599, 1997.

[Schultz, 1986] W. Schultz. Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of Neurophysiology*, 56 :1439–1462, 1986.

[Seamans *et al.*, 1995] J. K. Seamans, S. B. Floresco, et A. G. Phillips. Functional differences between the prelimbic and anterior cingulate regions of the rat prefrontal cortex. *Behavioural neuroscience*, 109(6) :1063–1073, 1995.

[Seamans et Phillips, 1994] J. K. Seamans et A. G. Phillips. Selective memory impairments produced by transient lidocaine-induced lesions of the nucleus accumbens in rats. *Behavioural neuroscience*, 108 :456–468, 1994.

[Seth, 1998] A. K. Seth. Evolving action selection and selective attention without actions, attention, or selection. In *From animals to animats 5 : Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, éditeurs R. Pfeifer, B. Blumberg, J.-A. Meyer, et S. W. Wilson, pages 139–147, Cambridge, MA, 1998. The MIT Press.

[Skinner, 1938] B. F. Skinner. *The behavior of organisms*. Appleon Century Croft, New York, 1938.

[Slotine et Lohmiller, 2001] J. J. E. Slotine et W. Lohmiller. Modularity, evolution, and the binding problem : a view from stability theory. *Neural networks*, 14(2) :137–145, 2001.

[Snaith et Holland, 1991] S. Snaith et O. Holland. An investigation of two mediation strategies suitable for behavioural control in animals and animats. In *From animals to animats : Proceedings of the First International Conference on Simulation of Adaptive Behavior*, éditeurs J.-A. Meyer et S. W. Wilson, pages 255–262, Cambridge, MA, 1991. MIT Press.

[Spier et McFarland, 1996] E. Spier et D. McFarland. A fine-grained motivational model of behaviour sequencing. In *From animals to animats 4 : Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, éditeurs P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack, et S. W. Wilson, pages 255–263, Cambridge, MA, 1996. The MIT Press.

[Srinivasan *et al.*, 1999] M. V. Srinivasan, J. S. Chahl, K. Weber, et S. Venkatesh. Robot navigation inspired by principles of insect vision. *Robotics and Autonomous Systems*, 26 :203–216, 1999.

[Suri et Schultz, 1999] R. E. Suri et W. Schultz. A neural network learns a spatial delayed response task with a dopamine-like reinforcement signal. *Neuroscience*, 91(3) :871–890, 1999.

[Suri et Schultz, 2001] R. E. Suri et W. Schultz. Temporal difference model reproduces anticipatory neural activity. *Neural Computation*, 13 :841–862, 2001.

[Sutton et Barto, 1998] R. S. Sutton et A. G. Barto. *Reinforcement Learning : An Introduction*. The MIT Press, Cambridge, MA, 1998.

[Tabuchi *et al.*, 2003] E. Tabuchi, A. B. Mulder, et S. I. Wiener. Reward value invariant place responses and reward site associated activity in hippocampal neurons of behaving rats. *Hipocampus*, 13(1) :117–132, 2003.

[Takanishi *et al.*, 1998] A. Takanishi, T. Aoki, M. Ito, et J. Yamaguchi Y. Ohkawa. Interaction between creature and robot-development of an experiment system for rat and rat robot interaction. In *IEEE international workshop on intelligent robots and systems (IROS'98)*, pages 1975–1980, 1998.

[Thierry *et al.*, 2000] A.-M. Thierry, Y. Gioanni, E. Dégénetais, et J. Glowinski. Hippocampo-prefrontal cortex pathway : anatomical and electrophysiological characteristics. *Hippocampus*, 10 :411–419, 2000.

[Tinbergen, 1951] N. Tinbergen. *The study of instinct*. Oxford University Press, London, 1951.

[Toates, 1986] F. Toates. *Motivational systems*. Cambridge University Press, Cambridge, 1986.

[Tolman et Honzik, 1930] E. C. Tolman et C. H. Honzik. Principles of purposive behavior. *University of Californie publications in psychology*, 4 :215–232, 1930.

[Trullier *et al.*, 1997] O. Trullier, S. Wiener, A. Berthoz, et J.-A. Meyer. Biologically-based artificial navigation systems : Review and prospects. *Progress in Neurobiology*, 51 :483–544, 1997.

[Trullier, 1998] O. Trullier. *Elaboration et traitement des représentations spatiales servant à la navigation chez le rat : Enregistrements électrophysiologiques et modèles*. PhD thesis, LIP6/AnimatLab, Université Paris 6, France, 1998.

[Tyrrell, 1993a] T. Tyrrell. *Computational mechanisms for action selection*. PhD thesis, Centre for Cognitive Science, University of Edinburgh, 1993.

[Tyrrell, 1993b] T. Tyrrell. The use of hierarchies for action selection. *Adaptive Behavior*, 1(4) :387–420, 1993.

[Webb et Consi, 2001] éditeurs B. Webb et T. R. Consi. *Biorobotics, methods and applications*. AAAI Press/MITPress, Cambridge, MA, 2001.

[Webb et Scutt, 2000] B. Webb et T. Scutt. A simple latency dependent spiking neuron model of cricket phonotaxis. *Biological Cybernetics*, 82 :247–269, 2000.

[Webb, 1995] B. Webb. Using robots to model animals : a cricket test. *Robotics and Autonomous Systems*, 16 :117–134, 1995.

[Webb, 2001] B. Webb. Can robots make good models of biological behaviour ? *Behavioral and Brain Sciences*, 24(6), 2001.

[Werner, 1994] G. Werner. Using second order neural connections for motivation of behavioural choice. In *From animals to animats 3 : Proceedings of the Third International Conference on the Simulation of Adaptive Behavior*, éditeurs D. Cliff, P. Husbandsand J.-A. Meyer, et S. W. Wilson, pages 154–164, Cambridge, MA, 1994. MIT Press.

[Wickens et Kötter, 1995] J. Wickens et R. Kötter. Cellular models of reinforcement. In *Models of Information Processing in the Basal Ganglia*, éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser, pages 187–214. The MIT Press, Cambridge, MA, 1995.

[Wickens, 1997] J. Wickens. Basal ganlia : Structure and computations. *Network : Computation in Neural Systems*, 8 :77–109, 1997.

[Wiepkema, 1971] P. R. Wiepkema. Positive feedback at work during feeding. *Behaviour*, 39 :266–273, 1971.

[Wilson, 1991] S. W. Wilson. The animat path to ai. In *From animals to animats : Proceedings of the First International Conference on Simulation of Adaptive Behavior*, pages 15–21, Cambridge, MA, 1991. The MIT Press/Bradford Books.

[Wood *et al.*, 2001] R. Wood, K. Gurney, et P. Redgrave. 'direct pathway' connextions to globus pallidus in a computational model of the basal ganglia. In *British neuroscience association abstracts*, volume 16, page 90, 2001.

[Woodward *et al.*, 1995] D. J. Woodward, A. B. Kirillov, C. D. Myre, et S. F. Sawyer. Neostriatal circuitry as a scalar memory : Modeling and ensemble neuron recording. In *Models of Information Processing in the Basal Ganglia*, éditeurs J. C. Houk, J. L. Davis, et D. G. Beiser, pages 315–336. The MIT Press, Cambridge, MA, 1995.

[Wu *et al.*, 2000] Y. Wu, S. Richard, et A. Parent. The organization of the striatal output system : a single-cell juxtacellular labeling study in the rat. *Neuroscience Research*, 38 :49–62, 2000.

[Yamada *et al.*, 1989] W. Yamada, C. Koch, et P. Adams. Multiple channels and calcium dynamics. In *Methods in neuronal modelling*, éditeurs C. Koch et C. Segev. The MIT Press, Cambridge, MA, 1989.

[Zahm et Brog, 1992] D. S. Zahm et J. S. Brog. Commentary : on the significance of the core-shell boundary in the rat nucleus accumbens. *Neuroscience*, 50 :751–767, 1992.