



HAL
open science

Contribution d'un modèle computationnel de sélection de stratégies de navigation aux hypothèses relatives à l'apprentissage spatial

Laurent Dollé

► **To cite this version:**

Laurent Dollé. Contribution d'un modèle computationnel de sélection de stratégies de navigation aux hypothèses relatives à l'apprentissage spatial. Automatique / Robotique. Université Pierre et Marie Curie - Paris VI, 2010. Français. NNT: . tel-00647199

HAL Id: tel-00647199

<https://theses.hal.science/tel-00647199>

Submitted on 1 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT DE L'UPMC- SORBONNE UNIVERSITÉS
ECOLE DOCTORALE INFORMATIQUE, TÉLÉCOMMUNICATION ET ELECTRONIQUE
DE PARIS

Spécialité:
Informatique

Présentée par:
Laurent DOLLÉ

Contribution d'un modèle computationnel de sélection de stratégies de navigation aux hypothèses relatives à l'apprentissage spatial

Soutenance le 29 octobre 2010 devant le jury composé de:

Dr. Ricardo CHAVARRIAGA (*EPFL, Lausanne*)
Pr. Bruno GAS (*UPMC, Paris*)
Pr. Philippe GAUSSIER (*Université de Cergy-Pontoise*)
Dr. Agnès GUILLOT (*UP Ouest, UPMC Paris*)
Dr. Etienne SAVE (*Université de Provence, Marseille*)
Dr. Denis SHEYNIKHOVICH (*UPMC, Paris*)

Co-directeur de thèse
Examineur
Rapporteur
Co-directeur de thèse
Rapporteur
Invité



Résumé

De nombreuses expériences montrent la capacité des mammifères, particulièrement étudiée chez le rat, à pouvoir utiliser plusieurs types de navigation pour rejoindre un but. Ces stratégies seraient utilisées en fonction de la disponibilité de différents types d'amers présents dans l'environnement, situés à la proximité ou loin du but (resp. amers proximaux et distaux), ainsi que d'autres facteurs, comme la motivation, l'expérience ou le stress.

Les études ont notamment porté ces dernières décennies sur les systèmes de mémoires engagés dans l'utilisation de ces stratégies, l'hypothèse dominante étant que des modules apprennent en parallèle des types distincts de stratégies, reposant ou non sur la construction d'une représentation interne de l'environnement. Les unes, reposant sur l'utilisation d'amers visibles, seraient apprises de manière relativement inflexible (apprentissage procédural) et impliqueraient notamment le striatum dorso-latéral. D'autres impliqueraient la formation hippocampique et reposeraient sur la construction préalable d'une "carte cognitive" de l'environnement (apprentissage déclaratif), en intégrant de manière redondante les amers disponibles. Cette carte, apprise indépendamment de la position du but, conférerait une flexibilité importante à ce type de stratégies puisqu'elle aurait la capacité de repositionner rapidement le but, si celui-ci venait à être déplacé. L'hypothèse d'une telle représentation spatiale (théorie dite "cognitive") et de systèmes parallèles ayant des capacités différentes d'apprentissage est supportée par des comparaisons entre animaux intacts et animaux ayant des lésions des structures nerveuses concernées, de même que des expériences impliquant des manipulations de l'environnement (e.g., déplacement d'amers).

Cette hypothèse est toutefois remise en question par la nature de l'apprentissage impliqué : la construction d'une représentation spatiale et son utilisation est contraire à nombre d'observations tendant à montrer que le comportement spatial est avant tout dirigé par une compétition entre les amers disponibles, résultant du même apprentissage procédural que celui mis en œuvre dans un conditionnement opérant (théorie dite "associative"). Des expériences démontrent en effet que certains amers peuvent être occultés ou bloqués par d'autres selon qu'ils sont appris parallèlement ou séquentiellement, et remettent ainsi en question la possibilité d'intégrer ces amers de manière redondante. Selon les tenants de cette théorie, la navigation serait donc avant tout apprise et gérée de manière unitaire, empêchant ainsi l'émergence de plusieurs stratégies apprises en parallèle. Nous adoptons ici l'hypothèse, déjà formulée par d'autres neurobiologistes, que ces deux points de vue antagonistes pourraient s'expliquer par une gestion modulaire de la navigation, permettant des apprentissages différenciés et parallèles de plusieurs stratégies, et dont le mécanisme de sélection pourrait favoriser soit les compétitions entre amers, soit la nécessité de leur intégration.

Cette thèse vise à concevoir un modèle computationnel de sélection de stratégies de navigation qui concilie les deux théories ou, a minima, y apporte des éléments de débats. La conception du modèle computationnel proposé dans ce travail et sous-tendu par cette hypothèse a nécessité de fait la mise en place d'une règle de sélection ne dépendant pas de la nature de l'apprentissage des stratégies et pouvant privilégier soit certains amers, soit l'utilisation d'une représentation spatiale construite avec plusieurs amers. Ce modèle sélectionne en ligne des stratégies apprises

par des algorithmes d'apprentissage procéduraux (apprentissage par renforcement) et déclaratifs (recherche de graphe). Ces acquisitions s'effectuent en parallèle et peuvent favoriser une coopération entre les stratégies, car chacune peut apprendre des comportements de l'autre. Le module de sélection de ces stratégies peut inciter des interactions compétitives, car il effectue ses choix sur la base d'un apprentissage de type associatif. L'existence d'un tel module de sélection, indépendant des systèmes d'apprentissage des stratégies, est corroboré par certains travaux récents démontrant l'implication de structures du cortex préfrontal et des ganglions de la base dans cette fonction.

Le dispositif expérimental utilisé est la piscine de Morris, dans laquelle le comportement spatial de rats a été étudié de façon approfondie depuis une trentaine d'années. Nous avons limité nos analyses, d'une part, aux interactions entre stratégies de guidage (utilisant directement des indices visuels) et stratégies de lieu (utilisant une représentation spatiale construite à partir de cellules de lieu simulées par un modèle d'hippocampe intégré au modèle) et, d'autre part, à la recherche d'un seul but, dépendant d'un seul système motivationnel.

Après avoir démontré le comportement du modèle dans des situations où seuls certains types d'amers (proximaux ou distaux) sont présents, nous avons reproduit diverses expériences dans lesquelles l'influence de l'interaction de ces types d'amers a donné lieu à des interprétations issues de l'une ou de l'autre des deux théories antagonistes. Nous y apportons un corpus d'explications, conciliant à la fois les données neurobiologiques et comportementales, ainsi que des prédictions comportementales qui permettront aux neurobiologistes d'estimer l'utilité de ce modèle. Nous avançons notamment que l'utilisation d'une stratégie fondée sur une carte cognitive (intégration d'amers) ne peut pas être écartée, car elle se révèle indispensable pour expliquer certains effets de blocage ou d'occultation caractéristiques de la théorie associative. Nous suggérons aussi que l'existence de deux stratégies de guidage, l'une ayant un cadre de référence égocentré (apprenant les trajets conduisant au but en fonction de l'orientation du corps), l'autre un cadre de référence allocentré (apprenant une direction générale conduisant au but en fonction d'un repère absolu) doit nécessairement être prise en compte pour générer certaines interactions compétitives ou coopératives observées entre systèmes de navigation. Nous montrons aussi qu'une intégration d'amers n'est pas forcément utilisée par les animaux car une stratégie de guidage allocentrée, même fondée sur une compétition d'amers, peut être suffisante dans certains contextes pour expliquer leur comportement.

En résumé, ce modèle de navigation, associant des systèmes parallèles d'apprentissage procéduraux et déclaratifs et sélectionnés par un mécanisme procédural, est en mesure de proposer une architecture computationnelle qui pourrait concilier les hypothèses issues des théories associative et cognitive du comportement spatial. Ce travail a également une retombée dans le domaine de la robotique, proposant par cette architecture une augmentation de l'autonomie dans la navigation d'un système artificiel, pouvant sélectionner en ligne les stratégies les plus efficaces pour atteindre ses ressources.

Abstract

A large number of experiments have shown the capacity of mammals, in particularly studies involving rats, to use several methods of navigation to reach a goal. These so-called "strategies" are chosen depending on the availability of different kinds of landmarks within the environment. These landmarks can be situated either close to or far from the goal (resp. proximal and distal landmarks). In the last few decades, studies have been particularly focused on the memory systems engaged in the use of these strategies, the dominant hypothesis being that systems learn in parallel distinct strategies that may rely or not on the construction of an internal representation of the environment. Some strategies –based on the use of visible landmarks– will be learnt in a relatively inflexible way (i.e. procedural learning) and presumably involve the dorsolateral striatum. Other strategies involve a hippocampus-dependant training phase and are based on the preliminary construction of a "cognitive map" of the environment (i.e. declarative learning), by integrating in a redundant manner the available landmarks. This map is learnt independently of the position of the goal providing more flexibility to this type of strategy since it would be able to quickly relocate the target, if it is moved to another location. The hypothesis of such a spatial representation (a theory known as "cognitive") and of parallel systems having different learning capacities is supported by comparisons between intact animals with lesioned animals of the neural structures concerned.

However this hypothesis can be reappraised with respect to the nature of the learning involved : the construction and use of a spatial representation is not supported by a number of observations showing that spatial behaviour is mainly steered by a competition between the available landmarks, thus suggesting the same procedural learning displayed in operating conditioning (a theory known as "associative"). Experiments show that certain landmarks can be overshadowed or blocked by others according to whether they were learned in parallel or sequentially, questioning the possibility of integrating these landmarks in a redundant manner. According to this theory, navigation would be above all learned and managed in a unitary manner, and thus prevent the emergence of several strategies learned in parallel.

We hypothesize that these two conflicting points of view can be explained by a modular framework of navigation, allowing differentiated and parallel learning of various strategies and a selection mechanism that could favor either the competition between landmarks, or the necessity of their integration.

We design a computational model which puts in place a rule of selection between strategies independent of their learning method and which can privilege either certain landmarks or the use of a spatial representation constructed with several landmarks. This model selects among strategies that are learnt online by algorithms of procedural learning (learning by reinforcement) and declarative learning (graph search). Learning of these strategies takes place in parallel and can promote a cooperation between the two strategies, as one strategy can learn the behaviours of the other. Alternatively, the selection mechanism can prompt competitive interactions, since its choices are based on an associative type of learning. The existence of such a selection module, independent of strategy learning systems, is supported by certain recent studies showing the

implication of prefrontal cortex structures and basal ganglia within this function.

We test the model by reproducing experiments performed in the Morris Water Maze, which has been extensively used for the last thirty years to study spatial behaviour of rats. We have limited our analyses to, on the one part, interactions between guidance strategies (directly using visual cues) and place strategies (using a cognitive map constructed from "hippocampal" place cells); and on the other part to the seeking of a single goal, dependent on a single motivational system.

After assessing the model behaviour in situations where only certain types of landmarks (proximal or distal) are present, we reproduced various experiments in which the mentioned opposing theories have different interpretations with respect to the nature of interaction of these types of landmarks. Our results provide a set of explanations, that reconcile both the neurobiological and behavioural observations. Moreover, we also provide behavioural predictions which will allow neurobiologists to appraise the usefulness of this model.

We propose notably that the use of a strategy based on a cognitive map (integration of landmarks) cannot be disregarded, as it shows itself to be required to explaining certain effects of blocking or overshadowing characteristic of the associative theory. We also suggest that the existence of two guidance strategies, one having an egocentric framework (learning routes leading to the target according to the orientation of the body), the other an allocentric framework (learning a general direction leading to the target according to an estimated absolute reference landmark) must be taken into account in order to generate competitive or cooperative interactions observed between systems of navigation. We also show that an integration of landmarks is not necessarily used by animals, as an allocentric guidance strategy, even based on a competition of landmarks, can be sufficient in certain contexts to explain their observed behaviour.

To summarise, the architecture of the proposed navigational model, including parallel systems of procedural learning and declarative learning selected through a procedural mechanism, is able to reconcile hypotheses issued from associative and cognitive theories of spatial behaviour. This study also contributed to the field of robotics since our model was able to control a navigating robot and to select in line the most effective strategies to reach available resources.

Remerciements

Durant ces quatre années, un certain nombre de personnes m'ont permis, par leur aide portée, leur discussion où simplement leur présence de mener ce travail à bien.

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 Philippe Gaussier et Etienne Save pour avoir accepté de relire ce manuscrit conséquent, leurs remarques pertinentes et leur présence à la soutenance malgré leurs nombreuses contraintes. Monsieur Bruno Gas pour avoir accepté de présider le Jury en dépit d'un emploi du temps surchargé.

J'aimerais remercier Agnès Guillot d'avoir accepté de co-diriger cette thèse avec une constance et une implication peu communes. Son support scientifique et moral, de même que sa participation active aux travaux présentés ici ainsi qu'à leur rédaction m'ont été d'une aide inestimable. Je remercie également Monsieur Ricardo Chavarriaga d'avoir co-dirigé cette thèse, pour ses nombreuses et constructives remarques, et bien sûr pour avoir fourni le modèle au cœur de cette thèse.

Je remercie Denis Sheynikhovich pour m'avoir grandement aidé durant trois courts mais intenses mois à développer le principal apport de cette thèse, les discussions fructueuses et son écoute toujours disponible même après son départ de notre équipe.

J'aimerais remercier les membres de l'équipe SIMA de l'ISIR pour leurs interactions, commentaires et participation non limitées au domaine scientifique : tout d'abord, Jean-Arcady Meyer pour avoir su réunir une équipe joviale et passionnée, Benoît Girard pour ses nombreux commentaires constructifs, Mehdi Khamassi pour les discussions passionnantes scientifiques, politiques et artistiques, Ken Caluwaerts pour l'incroyable travail réalisé sur le portage robotique du modèle, Jean Liénard pour avoir montré sa "biomimétisation" possible, Stéphane Doncieux, Jean-Baptiste Mouret, Sylvain Koos, Paul Tonelli et Tony Pinville pour les divertissantes et souvent passionnantes discussions lors des repas de midi. Je ne saurais oublier mes deux compagnons de thèse, Steve Nguyen et Louis-Emmanuel Martinet pour avoir enduré en même temps les mêmes épreuves, les mêmes doutes et les mêmes soulagements pendant ces quatre années.

Enfin, je remercie ma famille (mes parents, mes deux frères) pour leur support affectif, mes colocataires (Karim, Kirsten, Manu) pour avoir enduré mes choix musicaux parfois peu consensuels lors des phases d'écritures dans le salon le week-end, et mes amis, trop nombreux pour être cités ici.

Table des matières

I	Problématique	1
1	Introduction	3
1.1	Cadre et objectif général	3
1.2	L'approche animat	4
1.3	La navigation autonome	4
1.3.1	Différentes stratégies, différentes structures nerveuses	5
1.3.2	Un débat ancien mais toujours vif	5
1.3.3	Contribution des modèles computationnels	6
1.4	Objectif spécifique de ce travail	7
1.4.1	Un modèle computationnel modulaire de sélection de stratégies	7
1.4.2	Cadre de son application	7
1.5	Plan du mémoire	8
II	États de l'art	9
2	Données biologiques	11
2.1	Les différents types d'informations disponibles	11
2.1.1	Informations idiothétiques	11
2.1.2	Informations allothétiques	12
2.1.3	Le cadre de référence de ces informations	13
2.2	Les stratégies de navigation utilisant ces informations et cadres de référence	13
2.2.1	Exploration	14
2.2.2	Stratégies de guidage, associées à la position relative d'un but	14
2.2.3	Stratégies de lieux, associées à la position absolue d'un but	16
2.2.4	Stratégies de routes	18
2.3	Deux conceptions de l'apprentissage spatial	18
2.3.1	Théorie associative et apprentissage spatial	18
2.3.2	Théorie cognitive et apprentissage spatial	22
2.4	Un apprentissage spatial dépendant de systèmes neuraux distincts	24
2.4.1	Des changements de stratégies.	25
2.4.2	Des fonctions différentes entre systèmes neuraux.	26
2.5	Interactions entre systèmes neuraux	28
2.5.1	Systèmes neuraux apprenant en parallèle, mais pas indépendants	28

2.5.2	Sélection entre systèmes	28
2.5.3	Nature compétitive et coopérative des interactions	29
2.6	Conclusion du chapitre	31
3	Modèles computationnels	33
3.1	Modèles d'interactions entre amers	33
3.1.1	Règle de Rescorla-Wagner (Rescorla et Wagner, 1972)	34
3.1.2	L'hypothèse du comparateur (Miller et Matzel, 1988)	35
3.1.3	Règle de sélection bayésienne	36
3.1.4	Avantages et limites des modèles d'interactions entre amers	36
3.2	Modèles de systèmes parallèles	37
3.2.1	Les différents apprentissages mis en œuvre	37
3.2.2	Modélisation computationnelle des stratégies	40
3.2.3	Modèles de sélection entre stratégies	45
3.2.4	Avantages et limites des modèles de sélection de stratégies	53
3.3	Conclusion du chapitre : vers un nouveau modèle de sélection de stratégies de navigation	53
III	Méthode	55
4	Description du modèle	57
4.1	Codage des experts	58
4.1.1	Deux experts de guidage procéduraux : les experts Taxon et Direction	58
4.1.2	Un expert de lieu procédural : l'expert Local	62
4.1.3	Un expert de lieu "déclaratif" : l'expert Planification	66
4.1.4	Expert Exploration	68
4.1.5	Apprentissages réciproques des stratégies	68
4.2	Module de sélection des experts	69
4.2.1	Critère de sélection	70
4.2.2	Modularité de la sélection	71
4.3	Analyse des données	72
IV	Résultats : Simulations de l'influence d'amers proximaux et/ou distaux sur l'apprentissage spatial	77
5	Présence d'un seul type d'amer	79
5.1	Présence des seuls amers proximaux	80
5.1.1	But visible fixe ou balisé (EXP1)	80
5.1.2	Un seul amer à côté d'un but invisible (EXP2)	82
5.1.3	Présence de deux amers à côté du but (McGregor et al, 2004a)	85
5.1.4	Bilan de la présence des seuls amers proximaux	91
5.2	Présence des seuls amers distaux	92
5.2.1	But invisible et fixe : expérience inspirée de celle de Morris et al (1982)	93

5.2.2	But invisible déplacé : expérience inspirée de celle de Steele et Morris (1999)	98
5.2.3	Bilan de la présence des seuls amers distaux	102
5.3	Conclusion du chapitre	103
6	Interactions entre amers proximaux ou/et distaux	107
6.1	Effet de gradient associatif du à la distance entre un amer proximal et un amer distal (Rodrigo et al, 2006)	107
6.1.1	Protocole et résultats observés	108
6.1.2	Protocole et résultats simulés	109
6.1.3	Discussion partielle	117
6.2	Effet de blocage d’amers distaux par un amer proximal dans l’expérience de Roberts et Pearce (1999)	117
6.2.1	Protocole et résultats observés	118
6.2.2	Protocole et résultats simulés	120
6.2.3	Discussion partielle	129
6.3	Absence d’occultation d’amers proximaux par les amers distaux (McGregor et al, 2004a)	130
6.3.1	Protocole et résultats observés	131
6.3.2	Protocole et résultats simulés	133
6.3.3	Discussion partielle	138
6.4	Conclusion du chapitre	141
7	Interactions entre systèmes de guidage et de lieu en situations de conflit	145
7.1	Compétition ponctuelle (Devan et White, 1999)	145
7.1.1	Protocole et résultats observés	146
7.1.2	Protocole et résultats simulés	146
7.1.3	Discussion partielle	154
7.2	Compétition graduelle (Pearce et al, 1998)	157
7.2.1	Protocole et résultats observés	157
7.2.2	Protocole et résultats simulés	159
7.2.3	Discussion partielle	166
7.3	Conclusion du chapitre	167
V	Discussion et conclusion	171
8	Discussion générale	173
8.1	Principes essentiels de l’architecture du modèle	174
8.2	Contribution des principes de la modélisation aux résultats acquis	176
8.2.1	Des experts variés et complémentaires	176
8.2.2	Apprentissages réciproques	177
8.2.3	Sélection associative adaptative	179
8.3	Contributions du modèle aux hypothèses relatives à l’apprentissage spatial	180
8.3.1	Apports généraux	180
8.3.2	Apports plus spécifiques du modèle pour les protocoles reproduits	185

8.4	Eléments limitants du modèle	188
8.4.1	Traitement sensoriel des amers	188
8.4.2	Mécanismes d'apprentissage simplifiés	188
8.4.3	Contraintes dans le réseau de sélection	191
8.5	Assignment de structures nerveuses aux experts et au module de sélection	192
8.6	Bilan : le modèle comme outil d'analyse et de prédiction	195
9	Conclusion et Perspectives	199
9.1	De <i>Rattus rattus</i>	200
9.2	... à <i>Psikharpax</i>	202
VI	Annexes	205
A	Table des figures	207
B	Paramètres du modèle	211
B.1	Paramètres modifiés selon ces expériences	212
B.1.1	Morris et al (1982)	212
B.1.2	Roberts et Pearce (1999)	212
B.1.3	Devan et White (1999)	212
B.1.4	Pearce et al (1998)	212
C	Expérience de Morris et al (1982)	213
D	Interactions entre amers proximaux ou/et distaux	215
D.1	Rodrigo et al (2006)	215
D.2	Expérience de Roberts et Pearce (1999)	217
D.2.1	Groupes avec Taxon	217
D.2.2	Groupe DL-sep	219
E	Interactions entre systèmes de guidage et de lieu en situation de conflit	221
E.1	Expérience de Devan et White (1999)	221
E.2	Expérience de Pearce et al (1998)	223
F	Expériences supplémentaires	225
F.1	Expérience de Chamizo et Rodrigo (2004) : limitation des stratégies de guidage	225
F.2	Expert Parois	227
F.3	Expérience de Hamilton et al (2008)	227
	Références	231

Abbreviations

Structures neurales

- DG** : Gyrus denté
- DLS** : Striatum dorsolatéral
- DMS** : Striatum dorsomédial
- EC** : Cortex enthorinal
- Hc** : Hippocampe
- mPFC** : Cortex Préfrontal médial
- OFC** : Cortex orbitofrontal
- PFC** : Cortex préfrontal
- PL/IL** : Prélimbique / Infralimbique
- STN** : Noyau subthalamique

Modélisation

- AC** : Cellule d'action
- TD-learning** : Apprentissage par différence temporelle
- PC** : Cellule de lieu
- PG** : Graphe de planification
- SC** : Cellule sensorielle
- S-R** : Stimulus Réponse

Groupes simulés (cf. Chapitre 4.3)

D : Direction (équivalents animaux Hc)

DL : Direction-Local

DP : Direction-Planification

L : Local (équivalents animaux DLS)

P : Planification (équivalents animaux DLS)

T : Taxon (équivalents animaux Hc)

TL : Taxon-Local

TP : Taxon-Planification

Première partie

Problématique

Chapitre 1

Introduction

1.1 Cadre et objectif général

Depuis un peu plus de vingt ans des modèles computationnels s’inscrivant dans le cadre de l’approche *animat* (Brooks, 1990; Meyer et Guillot, 1991; Wilson, 1991; Guillot et Meyer, 2000) conçoivent des mécanismes susceptibles à la fois d’améliorer l’autonomie des systèmes artificiels et de contribuer à une meilleure compréhension du comportement animal. Les mécanismes que ces modèles développent s’inspirent explicitement des remarquables capacités d’adaptation de différentes espèces animales.

La navigation spatiale animale a inspiré dans ce domaine un très grand nombre de modèles computationnels (e.g., pour revues, Trullier et al, 1997; Franz et Mallot, 2000; Filliat et Meyer, 2003; Meyer et Filliat, 2003; Arleo et Rondi-Reig, 2007), car c’est un exemple privilégié en matière d’autonomie pour rechercher et atteindre des ressources dans un environnement inconnu – capacité dont les robots actuels ne sont pas encore pourvus de manière satisfaisante. Par ailleurs, si depuis près d’un siècle la navigation a fait l’objet de très nombreuses expériences animales, en particulier chez le rat (e.g., Skinner, 1938; Tolman, 1948), il reste que certains mécanismes propres à l’apprentissage spatial animal ne sont pas encore élucidés.

Les modèles de navigation bioinspirés peuvent donc encore contribuer dans le domaine de la navigation au double objectif de l’approche *animat*. Nous proposons dans ce travail un nouveau modèle computationnel qui vise particulièrement à analyser l’influence des amers environnementaux sur l’apprentissage de divers types de navigation utilisés pour rejoindre un but. Ce thème est en effet encore l’objet de débats dans la communauté des biologistes (e.g., Stahlman et Blaisdell, 2009) et est encore trop peu investigué dans la communauté des roboticiens (Save et al, 2008).

Nous rappellerons d’abord brièvement le cadre de l’approche *animat*, puis nous exposerons la problématique spécifique développée dans ce travail, avant de décrire la structure de ce document.

1.2 L'approche animat

La philosophie sous-tendant l'approche animat est fondée sur l'idée que les capacités adaptatives, des plus simples (e.g., un réflexe) aux plus complexes (e.g., raisonnement humain) ne peuvent être détachées de leur enveloppe physique et des interactions de celles-ci avec le monde réel. Cette approche émerge au domaine de l'intelligence artificielle *située*, pour lequel les capacités adaptatives d'un système artificiel ne peuvent se développer que si ce système acquiert et traite les informations environnementales de façon autonome avec ses propres boucles sensori-motrices.

D'après Arkin (1998), l'inspiration du comportement des animaux s'est avérée féconde pour les systèmes artificiels : ainsi les différentes écoles psychologiques – behavioriste, cognitiviste – ont-elles donné l'idée de développer des systèmes sensori-moteurs simples ou des systèmes capables de planification. On ajoutera les écoles de psychologie écologique et systémique, qui ont inspiré l'aspect situé des travaux. Les systèmes vivants ont ainsi démontré leur capacité à apporter une contribution significative à la robotique autonome.

En retour, les systèmes artificiels ont pu aussi permettre de mieux comprendre le comportement animal. Webb (2001, 2009) a passé en revue un certain nombre d'exemples dans lesquels l'exécution robotique de systèmes inspirés des animaux, à des niveaux plus ou moins abstraits, ont permis la confirmation ou l'infirmité d'hypothèses biologiques, la complétion d'hypothèses existantes, voire la proposition d'hypothèses. A titre d'exemples, on peut citer l'évaluation des conditionnements classique et opérant et des bases neurales de l'apprentissage associatif (Saksida et al, 1997; Verschure, 1998; Dampier et al, 2000), l'évaluation de l'influence de l'environnement sur le comportement, comme le suivi de polarisation de la lumière par les fourmis du désert (Lambrinos et al, 1997) ou la proposition de cellules de transitions entre lieux dans l'hippocampe (Gaussier et al, 2007). Implémentée sur une plateforme robotique, l'incarnation physique des modèles animats permet de surcroît une meilleure prise en compte des environnements réels, dont nombre de facteurs potentiellement influents sont sous-évalués, voire ignorés en simulation.

En collaboration avec les biologistes, des aller-retours successifs entre les prédictions des simulations et leur confirmation par les expériences animales permettent d'affiner incrémentalement les modèles pour ensuite mieux expliquer les données réelles et les rendre plus efficaces pour l'autonomie des systèmes artificiels.

Notre modèle computationnel vise, d'une part, à expliquer les résultats d'expériences de navigation spatiale du rat – animal dont les mécanismes sont les plus investigués dans ce domaine. D'autre part, afin de rendre toujours possible son implantation robotique, il reste à un niveau élevé de modélisation et ne vise pas à reproduire les activations fines de neurones particuliers.

1.3 La navigation autonome

Les animaux montrent de grandes capacités de navigation en ayant à leur disposition plusieurs moyens de rejoindre des ressources, que ce soit pour s'accommoder d'environnements dynamiques ou inconnus, ou pour maximiser leur chance de survie, tout en restant dans leur

zone de viabilité ou de confort (i.e., l'espace des valeurs que peuvent prendre un certain nombre de variables essentielles pour rester en vie, Ashby, 1952). Ils peuvent donc servir de modèles à la conception et l'amélioration des systèmes de navigation autonomes et adaptatifs.

1.3.1 Différentes stratégies, différentes structures nerveuses

Ces moyens divers d'orientation dont les animaux disposent ont été abondamment étudiés en neuroscience. De très nombreuses expériences montrent la capacité des mammifères, particulièrement étudiée chez le rat, à pouvoir utiliser plusieurs types de navigation pour rejoindre un but. Les taxonomies les plus courantes de ces différentes "stratégies" les classent en fonction des informations qu'elles utilisent (idiothétiques, i.e., générées par le mouvement propre de l'animal, et/ou allothétiques, i.e., générées par des éléments extérieurs à l'animal), de leurs cadres de référence (égocentré, dont les mouvements sont centrés sur la position de l'animal ou allocentré, dont les mouvements sont centrés sur la position d'un élément du monde extérieur), des caractéristiques de leurs apprentissages (lent ou rapide, flexible ou non au changement de tâche) (Trullier et al, 1997; Franz et Mallot, 2000; Arleo et Rondi-Reig, 2007).

Des expériences d'inactivation ont impliqué différentes structures nerveuses dans l'apprentissage et l'exécution de ces différentes stratégies, notamment le striatum dorsolatéral (DLS) pour celles plutôt peu flexibles et égocentrées, et le système hippocampique, parahippocampique et le cortex medial préfrontal pour d'autres, plus flexibles et allocentrées (e.g., White, 2004, 2005).

Cette possibilité d'utiliser des stratégies diverses renforcerait la robustesse de la navigation des rats, leur donnant l'opportunité de rejoindre un but visible ou non, indicé ou non indicé par différents éléments caractéristiques de l'environnement plus ou moins proches de ce but (amers proximaux et distaux). Selon ces conditions environnementales et la disponibilité correspondante des informations allothétiques ou idiothétiques, les animaux semblent capables de choisir les stratégies les plus adéquates, d'en changer lorsque la disponibilité de l'une ou l'autre de ces informations change, ou de gérer les conflits qu'elles peuvent générer (e.g., Tamara et al, 2010). Des structures nerveuses auraient été aussi impliquées dans ces choix (e.g., Ragozzino et al, 1999; Rich et Shapiro, 2009). C'est cette capacité à sélectionner entre différents moyens de rejoindre un but qui est cruciale pour expliquer l'autonomie des mammifères dans la navigation.

1.3.2 Un débat ancien mais toujours vif

Alors que la nature des informations environnementales ou internes disponibles, des cadres de référence utilisés et des principales structures nerveuses impliquées dans les stratégies de navigation suscitent encore des investigations mais pas de contestations aiguës, il n'en est pas de même de la nature des mécanismes attribués aux apprentissages de ces stratégies. Les prémisses de ce débat remontent à un siècle (e.g., Skinner, 1938; Tolman, 1948), mais il s'avère toujours actuel (e.g., Pearce, 2009; Leising et Blaisdell, 2009). Il met en présence, d'un côté, les partisans d'un apprentissage spatial répondant aux mêmes lois que celles du conditionnement opérant, sous la dépendance d'un mécanisme général fondé sur une *compétition* d'amers (théorie associative),

de l'autre les partisans d'un apprentissage spatial capable *aussi* de se fonder sur une *intégration coopérative* des amers disponibles (*cognitive mapping theory*, que nous nommerons théorie cognitive). En d'autres termes, l'apprentissage spatial ne serait dépendant que d'un système de mémoire procédural (i.e., encoder "comment faire" pour rejoindre un but de façon stéréotypée) ou pourrait aussi dépendre d'un système de mémoire déclaratif (i.e., encoder "quoi faire" pour rejoindre un but, indépendamment de ce but).

Certaines expériences, apportant des démonstrations convaincantes, contribuent à renforcer la première hypothèse (e.g., Chamizo, 2003; Leising et Blaisdell, 2009; Stahlman et Blaisdell, 2009) tandis que d'autres appuient la seconde (e.g., Tolman, 1948; Steele et Morris, 1999). Les principaux points d'achoppement portent sur l'existence d'une carte "cognitive", représentation spatiale dont les règles d'apprentissage sont contraires à la compétition d'amer (e.g., O'Keefe et Nadel, 1978) et, plus récemment, d'un module géométrique, dont l'utilisation est dépendante de la géométrie de l'environnement, qui serait "impénétrable" pour certains aux effets compétitifs (e.g., Cheng, 1986; Pearce et al, 2006; Redhead et Hamilton, 2007).

Si ces expériences comportementales n'ont jusqu'à maintenant pas réussi à clore définitivement ce débat, elles ont néanmoins fourni un nombre très conséquent d'informations sur les différentes natures possibles de l'apprentissage spatial.

1.3.3 Contribution des modèles computationnels

La notion de stratégies de navigation variées a été largement exploitée par les informaticiens et roboticiens, modélisées au moyen d'apprentissages procéduraux et déclaratifs. Cependant la grande majorité des modèles computationnels bio-inspirés ne se sont concentrés que sur la modélisation de chaque type de navigation (e.g., pour revue, Filliat et Meyer, 2003; Meyer et Filliat, 2003), plutôt que sur la façon dont elles pouvaient être choisies en fonction des conditions environnementales. Ces modèles peuvent de fait s'avérer fragiles lorsque l'environnement change et/ou que le type d'entrée utilisé auparavant s'avère inutilisable. En cela, ils privent les systèmes artificiels d'une augmentation d'autonomie conséquente en environnement inconnu.

A part une catégorie de modèles d'apprentissage spatial s'inscrivant dans la théorie associative (e.g., Rescorla et Wagner, 1972; Miller et Shettleworth, 2007; Stout et Miller, 2007), mais dont la nature formelle et analytique les prive d'une réelle assise bio-inspirée, aucun modèle – à notre connaissance – n'a tenté de s'interroger sur le déterminisme des différents résultats apportés par les deux écoles. Cela est cependant directement lié à la question de savoir comment des stratégies différentes peuvent être sélectionnées, par une compétition d'amers, par une intégration d'amers, ou les deux.

1.4 Objectif spécifique de ce travail

1.4.1 Un modèle computationnel modulaire de sélection de stratégies

Notre objectif est donc de proposer un modèle computationnel bio-inspiré – en accord avec la fonction de structures nerveuses impliquées dans la navigation spatiale –, qui pourrait gérer une sélection de diverses stratégies sous-tendues par des apprentissages de même nature ou non, de type procédural ou déclaratif, en fonction des informations disponibles dans l’environnement. Suivant les propositions de certains neurobiologistes (e.g., White, 2004, 2005) et s’inspirant d’un modèle déjà validé (Chavarriaga et al, 2005), il met en œuvre différents "experts" gérant l’apprentissage de différentes stratégies, ainsi qu’un module de sélection séparé qui estime à tout moment l’adéquation de chaque expert en fonction des conditions environnementales. Ce module de sélection, qui suit une règle associative, est capable d’estimer la validité ou l’efficacité de stratégies qui, elles, sont acquises selon un mode procédural ou déclaratif. Cela a nécessité la recherche non triviale d’une "monnaie commune" permettant la comparaison des différentes stratégies, quels que soient leurs modes d’acquisition.

En d’autres termes, ce modèle tente donc de concilier les résultats émanant de ces deux écoles et parfois contradictoires par une gestion modulaire de la navigation, permettant des apprentissages différenciés et parallèles de plusieurs stratégies, et dont le mécanisme de sélection pourrait favoriser soit les compétitions entre amers, soit la nécessité de leur intégration. Ce travail ne prétend évidemment pas apporter de réponse définitive au débat sur l’apprentissage spatial, mais vise à y apporter une première contribution.

1.4.2 Cadre de son application

Le domaine à explorer étant très vaste, nous le restreignons à certaines conditions. Nous ne nous intéresserons qu’aux expériences de navigation ne comportant qu’un seul but à atteindre, dépendantes d’une seule motivation. L’environnement modélisé sera la piscine de Morris (Morris, 1981) : cet environnement présente l’avantage d’avoir été utilisé de façon intensive dans les expériences de navigation depuis une trentaine d’années, de permettre un contrôle non négligeable sur l’influence des amers (la présence d’eau empêche notamment les indices olfactifs indésirables d’interférer) et n’est pas basée sur une déplétion énergétique des animaux, restreignant là encore l’éventuel rôle de diverses motivations sur la navigation (D’Hooge et De Deyn, 2001; Carillo-Mora et al, 2009).

Nos simulations se limiteront également aux interactions entre stratégies dites de guidage visuel (utilisant directement différents amers visuels) et de lieu (utilisant une intégration d’amers visuels). Les stratégies utilisant des informations purement idiothétiques ou dépendantes d’une configuration géométrique seront évoquées dans l’état de l’art mais ne seront pas prises en compte dans les simulations (ces dernières étant notamment au cœur d’un débat concernant l’existence d’un module géométrique dédié, e.g., Sheynikhovich et al, 2009).

1.5 Plan du mémoire

Nous aborderons dans les deux chapitres suivants (Chapitres 2 et 3) les états de l'art respectifs des données biologiques et des modèles computationnels liés à la problématique évoquée ci-dessus.

Dans les Chapitres 4 et 5 nous détaillerons la conception du modèle de sélection de stratégie de navigation ainsi que la méthodologie générale appliquée à l'analyse des résultats.

Les Chapitres 6 à 8 présenteront les résultats quant aux analyses d'interactions de stratégies dans des expériences émergeant ou non à l'une des deux théories associative ou cognitive :

(i) en présence d'un type d'amers seulement (soit proximaux, soit distaux : deux expériences simulées, puis expériences de McGregor et al, 2004a; Morris et al, 1982; Steele et Morris, 1999),

(ii) en présence des deux types d'amers (proximaux et distaux), dans des conditions faisant apparaître ou non des effets de l'apprentissage associatif (Rodrigo et al, 2006; Roberts et Pearce, 1999; McGregor et al, 2004a),

(iii) pour finir par deux expériences où les deux types de stratégies de guidage et de lieu sont mises délibérément en conflit (Devan et White, 1999; Pearce et al, 1998).

Le Chapitre 9 discutera des contributions et limitations du modèle quant aux hypothèses sur l'apprentissage spatial. Enfin le Chapitre 10 conclura sur le travail effectué et apportera des perspectives, notamment en matière d'implémentation de ce modèle sur un robot réel, le robot-rat Psikharpax.

Deuxième partie

États de l'art

Chapitre 2

Données biologiques

Ce chapitre se restreindra à exposer quelques données expérimentales issues des neurosciences essentielles pour comprendre la démarche de ce travail. Ces données porteront principalement sur le rat, qui est l'animal sur lequel l'apprentissage spatial a été le plus étudié, que ce soit du point de vue de son comportement ou de données neurobiologiques associées. Ce chapitre est divisé en quatre parties : les principaux facteurs sur lesquels se fonde l'apprentissage spatial (informations disponibles, cadre de référence), les différentes stratégies s'appuyant sur ces facteurs, les deux conceptions sur leurs apprentissages, associatif et cognitif, et enfin les différents systèmes neuraux supposés être associés à la gestion de ces apprentissages et leurs interactions.

2.1 Les différents types d'informations disponibles

Nous introduisons dans cette section les notions qui sont intervenues dans les classifications classiques des différentes stratégies de navigation, soit le type d'informations disponibles et leur cadre de référence. Nous débattons de leur type d'apprentissage dans une section suivante.

Les animaux utilisent les informations provenant de leur propre déplacement (informations idiothétiques) et des indices environnementaux (informations allothétiques) pour se localiser et retrouver un but. Elles sont utilisées de façon complémentaire par l'animal, comme le montrent les études manipulant respectivement les deux types d'informations (Stuchlik, 2002).

2.1.1 Informations idiothétiques

Les informations idiothétiques sont générées par l'animal lui-même et le renseignent sur son état interne pendant la navigation. Elles sont fournies par ses capteurs proprioceptifs, vestibulaires et kinesthésiques ou le flux optique (e.g., Watson, 1907; Barlow, 1964). Elles peuvent donc être présentes en dehors de tout indice environnemental particulier. Ces informations permettent à l'animal d'estimer son orientation et la distance parcourue par son mouvement propre (Etienne et Jeffery, 2004).

2.1.2 Informations allothétiques

Les informations allothétiques sont fournies par le milieu environnemental et consistent en des indices de différentes modalités sensorielles (visuelles, olfactives, auditives, etc., e.g., Lavenex et Schenk, 1998) qui renseignent l'animal sur l'état de l'environnement. Ce sont les indices visuels qui sont le plus fréquemment étudiés. Ces repères existent bien évidemment dans un environnement naturel, mais nous les classerons ici en fonction de leur disposition à l'intérieur ou à l'extérieur du dispositif expérimental.

Indices faisant partie du dispositif expérimental

Les amers proximaux. Ce sont des repères discrets qui sont disposés sur ou à côté du but et permettent à l'animal de s'en approcher et de tourner autour. Ils fournissent ainsi un aspect en 3D à l'animal (Save et al, 2008).

En fonction de la proximité du but, Gould et al (2010) distinguent :

(i) la balise (*beacon*) : lorsque l'amer est situé sur le but ; il suffit alors pour l'animal de s'en approcher pour atteindre le but

(ii) l'amer proximal (à proprement parler) : lorsque l'indice est situé à une distance raisonnable du but (< 1 m). De par sa position, l'animal peut interagir avec l'amer (i.e., tourner autour de lui), et estimer sa position par rapport à lui.

La configuration géométrique du dispositif. Des amers discrets ne sont toutefois pas les seuls indices permettant de trouver le but. Certaines expériences ont mis en évidence la possibilité pour les animaux de repérer le but à l'aide de la géométrie du dispositif, en se basant prioritairement sur les différents angles aigus ou obtus de cet environnement, tout en excluant les propriétés non géométriques (couleur, luminosité, etc., Cheng, 1986).

Les bords du dispositif. D'autres expériences ont également mis en évidence l'importance des bords d'un dispositif, repères non discrets aussi, sans forcément que ces bords forment un modèle géométrique bien défini : par exemple, les parois circulaires de la piscine de Morris ne permettent pas explicitement de différencier les lieux (e.g., Hamilton et al, 2008).

Indices ne faisant pas partie du dispositif expérimental : les amers distaux.

Selon Gould et al (2010), les amers distaux peuvent être à plusieurs mètres et cet éloignement les rendent plus stables pour l'animal que les amers proximaux. Ce sont des indices qui sont disposés dans la pièce où se trouve le dispositif ou sur des rideaux qui entourent le dispositif et sur lesquels sont accrochés différents repères. Ils ne peuvent pas être approchés par l'animal et ne lui fournissent donc qu'un aspect en 2D (Save et al, 2008).

Il faut noter toutefois que la distinction entre indices situés dans et en dehors du dispositif ne correspond pas toujours très clairement avec la distinction entre amers proximaux et distaux. Par exemple, des amers situés sur la paroi d'une piscine de Morris ont un statut particulier : ils appartiennent au dispositif mais ne peuvent pas être contournés par l'animal (e.g., Chamizo et Rodrigo, 2004). Il en est de même pour la configuration géométrique.

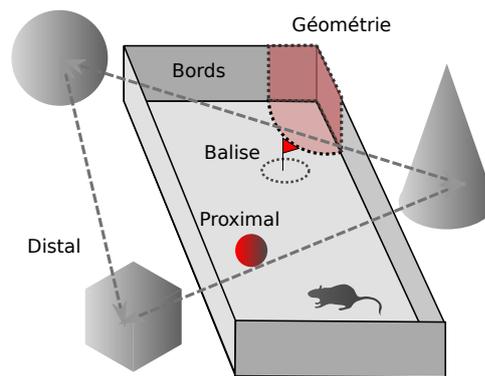


Figure 2.1 – Catégorisation des types d'amers et indices environnementaux permettant de localiser un but. Ce but se trouve ici sous la balise.

2.1.3 Le cadre de référence de ces informations

Les informations idiothétiques et allothétiques utilisées par les animaux pour l'apprentissage spatial peuvent être encodées dans deux types de référentiels (Berthoz, 1991; Klatzky, 1998, Figure 2.2). Un référentiel est dit égocentrique lorsque l'animal est son propre point de référence. La position des objets de son environnement est donc estimée par rapport aux propres position et orientation de l'animal (e.g., à droite, devant, etc).

Un référentiel allocentrique, en revanche, nécessite de l'animal l'utilisation d'un système de coordonnées indépendant de sa position et de son orientation. Il s'ancre sur un objet ou une direction cardinale pour orienter ses mouvements.

Une même information, idiothétique ou allothétique, peut-être encodée dans un référentiel égocentrique ou allocentrique (Burguière, 2006).

2.2 Les stratégies de navigation utilisant ces informations et cadres de référence

Nous avons choisi de distinguer dans cette section les stratégies avec lesquelles l'animal rejoint une position *relative* du but (lorsque le but est fixe et que le dispositif est déplacé, l'animal positionne le but à son emplacement supposé déplacé) et d'autres, avec lesquelles l'animal rejoint une position *absolue* du but (lorsque le but est fixe et que le dispositif est déplacé, l'animal positionne le but à son emplacement réel, Figure 2.3). Par simplification, nous nommerons les premières "*stratégies de guidage*" (bien que ce terme soit souvent associé à ce que nous nommons guidance dans le paragraphe suivant) et les secondes "*stratégies de lieu*". D'autres choix de taxonomies sont disponibles dans la littérature (e.g., Trullier et al, 1997; Squire et al, 2004; Sutherland et Hamilton, 2004; Arleo et Rondi-Reig, 2007).

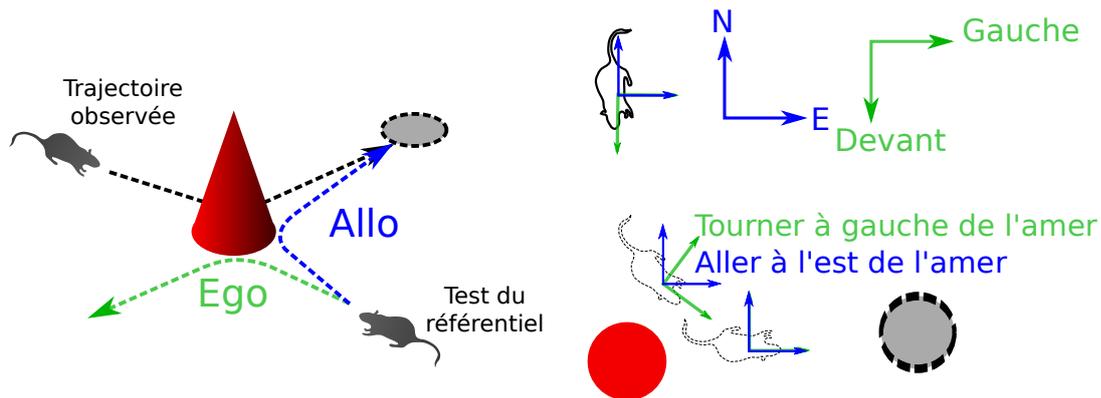


Figure 2.2 – Les deux référentiels utilisés par les animaux. Lors de la navigation de l’animal, les coordonnées égocentriques changent, à l’inverse des coordonnées allocentriques (inspiré de Klatzky, 1998). Par exemple (à gauche), dans l’expérience de Packard et McGaugh (1996), des rats sont entraînés à rejoindre un bras d’un T-maze en partant du même point de départ ("Trajectoire observée"). Si l’on fait partir les rats du point de départ opposé ("Test du référentiel"), ils iront dans un bras différent selon qu’ils ont appris à se déplacer selon un référentiel allocentrique (à l’Est) ou égocentrique (à leur gauche). A droite, on peut voir que le référentiel allocentrique est indépendant de l’orientation du rat, mais pas le référentiel égocentrique, au fur et à mesure que le rat approche le but.

2.2.1 Exploration

Paradoxalement nous commençons à citer une stratégie qui ne peut se ranger dans aucune de ces catégories. L’exploration est une stratégie de navigation qui ne peut en effet être classée ainsi mais qui soutient toujours tous les autres moyens de naviguer. Elle est fréquemment supposée aléatoire, mais peut relever aussi de comportements beaucoup plus déterministes (e.g., Eilam et Golani, 1989; Draï et al, 2001).

2.2.2 Stratégies de guidage, associées à la position relative d’un but

Praxis. La praxis est une stratégie égocentrée qui correspond à une suite ordonnée de mouvements de l’animal exécutés sur la base de ses informations idiothétiques. Ces informations consistent en une estimation de la vitesse de déplacement et de rotation. Elles permettent à l’animal d’estimer la distance parcourue depuis le point de départ pour connaître le prochain mouvement à effectuer. Une très ancienne expérience testant des rats sans vision, audition, odorat ni toucher dans un labyrinthe complexe, a prouvé qu’ils pouvaient réussir à retrouver de la nourriture (Watson, 1907). Le retour au nid (Maaswinkel et Whishaw, 1999, Figure 2.4 a) est un exemple de praxis, qui peut s’appuyer aussi sur des informations allothétiques : les animaux peuvent atteindre un but donné en suivant une trajectoire alambiquée (par exemple, en explorant l’environnement), mais reviennent à leur point de départ en ligne droite.

Cette stratégie a été citée aussi pour le parcours d’un T-Maze (Packard et McGaugh, 1996, Figure 2.4 b), où un rat apprenait à toujours réaliser le même virage. Après beaucoup d’essais,

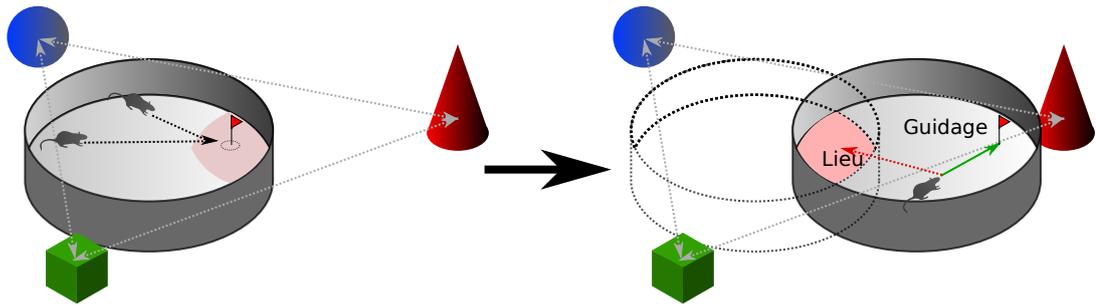


Figure 2.3 – Notre distinction entre une stratégie de "guidage" et une stratégie de "lieu", par rapport à la position relative ou absolue du but (voir définitions dans le texte).

le rat continuait ce comportement, même en l'absence de but. Cependant il est possible que l'animal se soit fixé sur des éléments géométriques du dispositif (coins, etc.).

Taxon (approche d'objet). Lorsque le but est visible ou signalé par une balise ou un amer proximal, les rats peuvent effectuer une stratégie d'approche d'objet, qui consiste à s'orienter en direction du but par des informations allothétiques et d'avancer jusqu'à l'atteindre (Figure 2.5 a, Collett et al, 1986; Biegler et Morris, 1993; Redhead et al, 1997; Timberlake et al, 2007; Waller et Lippa, 2007). Lorsque l'amer n'est pas une balise, l'animal atteint le but par exploration. Cette stratégie est exécutée dans un référentiel égocentrique, l'animal ayant juste à estimer sa position par rapport au but.

Guidance (nommé souvent guidage). Lorsqu'un ou des amers se trouvent moins proches d'un but caché, mais que ce but est situé en relation constante avec une configuration de ces amers, il est possible pour l'animal de le repérer grâce à la relation qu'il tient avec ces amers. L'animal pourra mémoriser la vue locale d'alors – un *snapshot* – et se déplacer de manière à minimiser la différence entre la vue courante et celle du *snapshot* obtenu à proximité du but (e.g., Collett, 1992, chez les insectes). Il atteint le but sans recherche exploratoire. Cette vue est égocentrique : l'animal doit se retrouver avec la même orientation par rapport au but pour que les deux *snapshots* correspondent (Figure 2.5b).

Direction Dans les mêmes circonstances que précédemment, l'animal peut mémoriser la direction générale à prendre pour rejoindre un but non visible en fonction de la position d'un ou de plusieurs amers proximaux ou distaux. A l'inverse de la stratégie Taxon, il ne recherche pas le but autour de l'amer par exploration, mais l'atteint relativement précisément. C'est l'équipe d'Hamilton et al (2008) qui a mis l'accent sur cette stratégie, qui semble être fréquemment utilisée par les rats. Pearce et al (1998) évoquent aussi la contribution d'informations idiothétiques, comme la notion de direction de la tête de l'animal (*Heading vector*). Elle serait utilisée dans un référentiel allocentrique. Ainsi l'animal apprend qu'un but est situé au nord d'un amer, ou à telle direction entre deux amers, quelle que soit son orientation (Pearce et al, 1998; Hamilton et al, 2008; Doeller et Burgess, 2008, Figure 2.5 c).

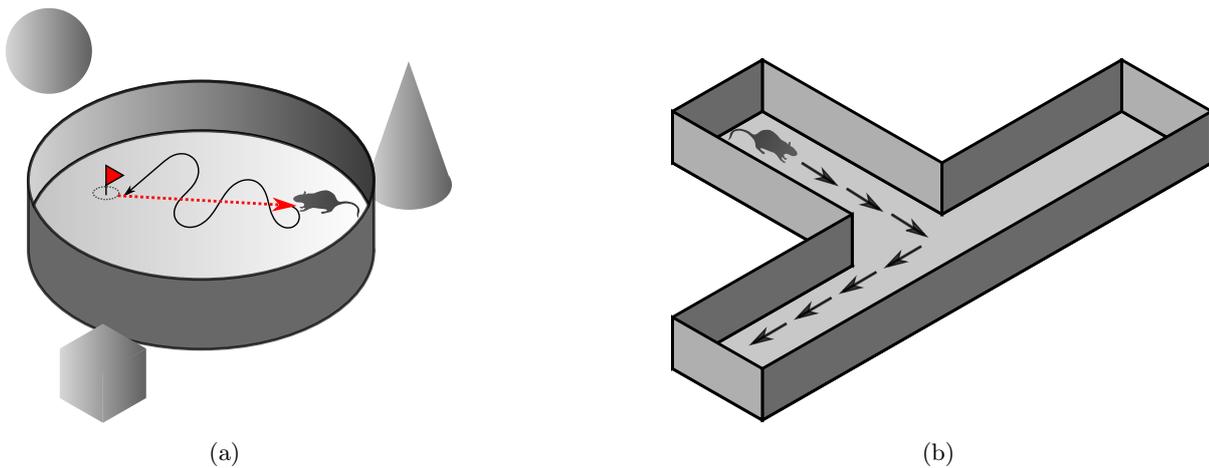


Figure 2.4 – (a) *Homing* ou retour au nid : Capacité à revenir vers le point de départ en ligne après avoir atteint un but par une trajectoire différente (par exemple, par exploration) (b) La stratégie Praxis mise en œuvre dans un T-Maze. Lorsqu'un rat apprend à toujours tourner à droite pour rejoindre un but, il persiste longtemps dans cette stratégie, même après disparition du but.

Stratégie géométrique et stratégie de bord. L'existence d'une stratégie de navigation basée sur la géométrie a été découverte par Cheng (1986). D'après Cheng (1986) et Gallistel (1990), l'animal extrait de la vision de l'environnement les informations purement géométriques, en excluant les autres informations (telles que couleur, luminosité, etc.). Ainsi l'emplacement du but est mémorisé par un petit nombre de paramètres, tels les principaux axes de l'environnement, et rejoindre le but consisterait à minimiser la différence entre les paramètres couramment extraits et ceux mémorisés.

Une stratégie encore mal investiguée, mais qui serait utilisée fréquemment, est celle qui consiste pour l'animal à positionner le but par rapport à une certaine distance des parois du dispositif (Hamilton et al, 2008). Elle s'appuierait donc sur des éléments non discrets du dispositif.

2.2.3 Stratégies de lieux, associées à la position absolue d'un but

Ces stratégies ont été établies par l'observation d'animaux capables de rejoindre un but caché situé dans un dispositif même lorsque celui-ci (mais pas le but) est déplacé.

Stratégie topologique. Contrairement à la stratégie Direction, une stratégie de lieu permet de localiser de façon absolue l'emplacement d'un but (i.e., par rapport à l'environnement général où se trouve l'animal) surtout grâce aux amers situés en dehors du dispositif (O'Keefe et Nadel, 1978). Cette stratégie est donc allocentrique.

L'animal pourrait acquérir une notion topologique de l'environnement, c'est-à-dire en ayant associé les transitions possibles entre différents lieux de l'environnement, sans pour autant avoir

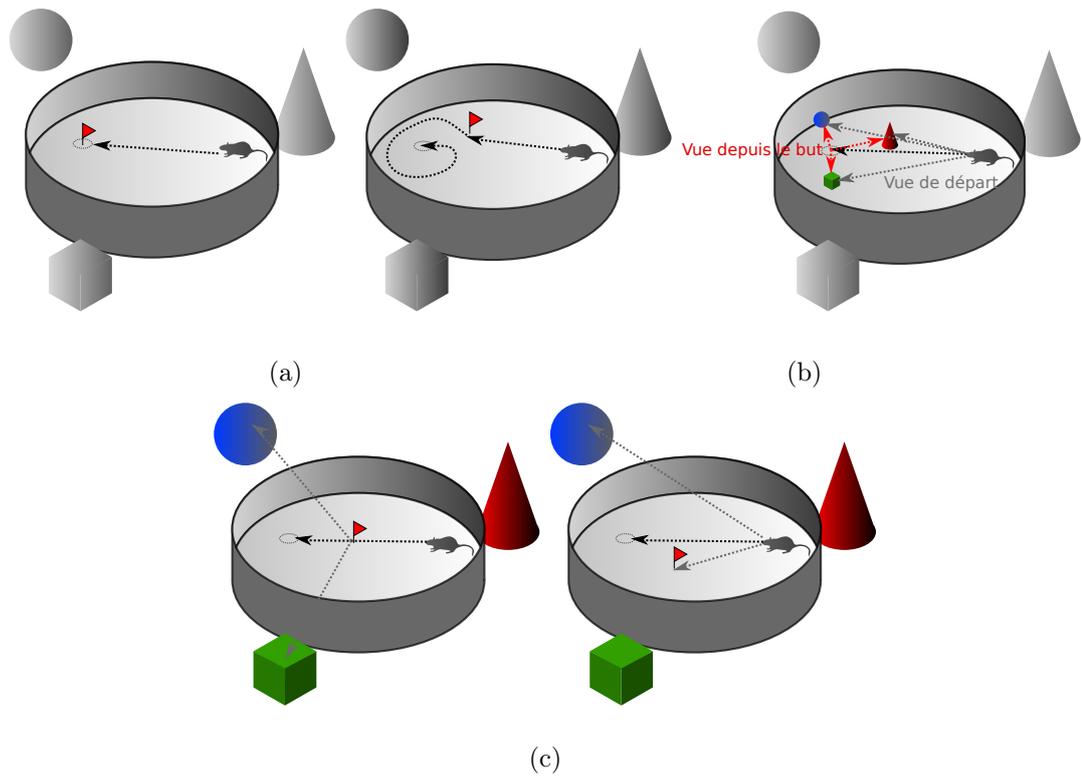


Figure 2.5 – (a-c) Principales stratégies de "guidage" : (a) Taxon : Le but est visible ou marqué par une balise (drapeau) : l'animal se dirige directement dans sa direction. Dans le cas où le but est indiqué, il le cherche à partir de l'amer. (b) Guidance : le but est caché mais sa position est mémorisée par la vue de l'environnement à cet endroit (flèches rouges), l'animal se déplace alors afin de minimiser la différence entre la vue courante (flèches grises) et celle mémorisée au but. (c) Direction : L'animal trouve le but caché de façon allocentrique par rapport à un ou plusieurs amers.

une notion précise de la distance entre ces lieux. Il pourra dans ce cas rejoindre le but par différents trajets, prendre des raccourcis ou contourner des obstacles par des chemins qu'il aura déjà explorés (e.g., Tolman, 1948; Morris, 1981).

Stratégie métrique. Une autre stratégie de lieu est une extension de la précédente, qui consiste à estimer la distance précise entre les lieux. Elle permettrait alors d'inférer des trajets nouveaux non encore explorés. Si cette stratégie a été avérée chez les humains et aussi chez les chiens (Chapuis, 1987; Thinus-Blanc, 1996), elle n'aurait pas encore été véritablement établie chez les rats (Poucet, 1993).

2.2.4 Stratégies de routes

Cette catégorie concerne les stratégies de guidage et de lieu, puisqu'il peut s'agir d'enchaînements séquentiels de stratégies de guidage ou de stratégies de lieu (Mittelstaedt et al, 1982; Rondi-Reig et al, 2006).

2.3 Deux conceptions de l'apprentissage spatial

Le débat déjà ancien (Skinner, 1938; Tolman, 1948) de la question de la nature de l'apprentissage de ces stratégies est toujours vif aujourd'hui (e.g., Kelly et Gibson, 2007; Leising et Blaisdell, 2009). Il oppose les partisans de mécanismes d'acquisition ne se distinguant pas de ceux sous-tendant les conditionnements classiques et opérants, à d'autres, qui évoquent un apprentissage spécifique au domaine spatial, dont les mécanismes seraient différents des précédents. Les premiers émargent à la théorie dite *associative*, les seconds à la théorie dite *cognitive mapping*, que nous nommons pour simplifier théorie cognitive. Le cas d'une conception consensuelle entre ces deux points de vue sera vue dans la prochaine section.

2.3.1 Théorie associative et apprentissage spatial

2.3.1.1 Une compétition entre les amers

Selon les partisans de cette théorie, l'apprentissage spatial serait sous la dépendance d'un seul type de mécanisme par lequel une nouvelle réponse est acquise par l'association d'un stimulus et d'une récompense (Hull, 1943). La caractéristique de cet apprentissage est qu'il se pratique incrémentalement, par essais et erreurs (Thorndike, 1911). Les mécanismes de ce type d'apprentissage ont été abondamment étudiés dans le cadre de conditionnements classique et opérant (Pavlov, 1927; Skinner, 1938; Watson, 1948). L'animal apprend donc, par exemple, l'emplacement d'un but en associant un amer (stimulus conditionné) avec le déplacement à effectuer (réponse) pour avoir une récompense.

Les mécanismes d'un tel apprentissage ont notamment été formalisés par Rescorla et Wagner (1972). La formulation mathématique de la règle correspondante sera développée plus précisément dans le Chapitre 3. L'idée sous-tendant cette règle est que les stimuli peuvent acquérir une force de prédiction de la récompense. Cette force est partagée entre tous les stimuli, si bien que lorsque l'un d'eux devient plus prédictif que d'autres, ces derniers ne peuvent obtenir

qu'une force de prédiction moins importante. Mackintosh (1975) nuancera toutefois cette vision en proposant plutôt que les associations entre stimuli sont indépendantes entre elles, mais que seule la plus forte d'entre elles est sélectionnée par un mécanisme d'attention pour diriger le comportement.

En d'autres termes, une des conséquences de cet apprentissage est que les différents stimuli ou groupes de stimuli disponibles dans l'environnement entrent en compétition pour contrôler le comportement de navigation de l'animal. Ceux qui ne sont pas favorisés par cette compétition ne vont pas contribuer à l'atteinte du but. Parmi les conséquences de cette compétition les plus étudiées sont les phénomènes dits de "blocage" (*blocking*) et d'"occultation" (*overshadowing*) (Kamin, 1969; Mackintosh, 1975).

Le *blocage* peut apparaître lorsque des stimuli sont appris *successivement*. Le blocage d'un stimulus B par rapport à un stimulus A s'observe lorsque l'animal est entraîné avec un premier stimulus A conditionné, qui est suivi d'une récompense inattendue, puis en présence de A et B, toujours suivi d'une récompense. L'animal va alors apprendre la réponse conditionnée par rapport à A, mais très peu, voire pas du tout, par rapport à B, ce que l'on peut tester avec une troisième étape en présence de B seulement, dans laquelle l'animal n'exprimera pas la réponse attendue (Figure 2.6 a).

L'*occultation* peut apparaître lorsque des stimuli sont appris *simultanément*. L'occultation du stimulus B par le stimulus A se produit lorsque les deux stimuli simultanément présents ne produisent pas la même association avec la réponse : le seul stimulus A, qui prédit mieux la récompense que B, prend le contrôle du comportement aux dépens du stimulus B. Si A est brusquement absent, la réponse ne peut être obtenue (Figure 2.6 b).

2.3.1.2 Expériences de navigation animale démontrant des effets "associatifs"

De nombreuses expériences comportementales soutiennent ces effets d'un apprentissage associatif, nous ne citerons que quelques exemples.

Gradient de distance au but. Certaines expériences semblent démontrer que la force d'association entre un amer et la réponse de l'animal dépend de sa distance au but. Plus un amer est proche du but, plus il contrôle le comportement de l'animal.

Rodrigo et al (2006) comparent la valeur prédictive de deux amers, un distal situé en dehors d'une piscine de Morris et un proximal situé sur la plateforme cachée. Après un entraînement en présence des deux amers, l'amer proximal est déplacé successivement à des distances de plus en plus éloignées de l'amer distal. Les rats expriment un gradient d'occupation de la zone autour de la balise proportionnel à la distance entre les deux amers. Les auteurs constatent également que ce gradient ne concerne pas la zone proche de l'amer distal, qui n'est plus que très peu occupée pendant les déplacements de l'amer proximal, ce qui montre que la balise a été plus influente que l'amer distal (qui a donc été occulté).

Avec trois amers proximaux, un situé au-dessus de la plateforme et deux situés sur la paroi de la piscine – l'un étant plus près du but que l'autre –, Chamizo et Rodrigo (2004) montrent

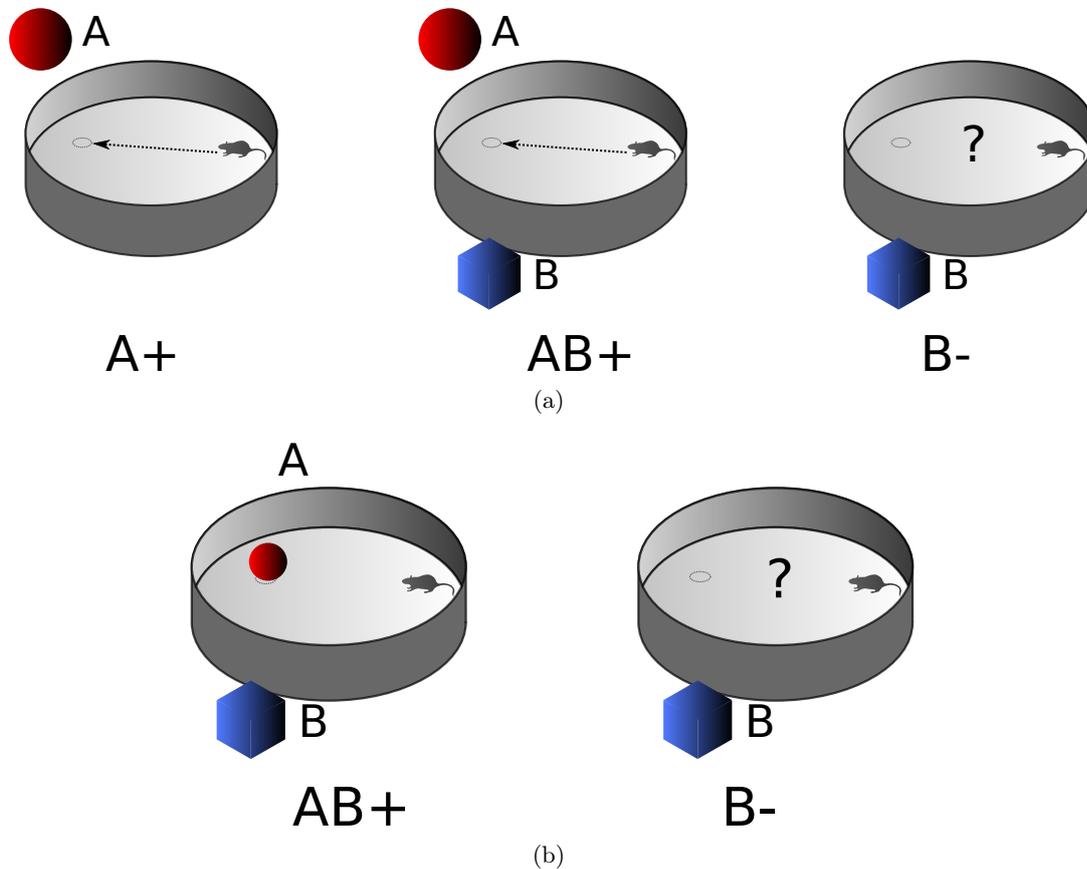


Figure 2.6 – a) Cas de blocage dans l'apprentissage spatial. L'animal apprend d'abord à associer la présence de l'amer A avec la présence de récompense (A+), puis l'on introduit l'amer B, toujours en présence de la récompense (AB+). Enfin si l'on retire A de l'environnement, et que l'on analyse le comportement du rat (B-), celui-ci ne montre aucune préférence en faveur de la position connue du but : Il n'a appris à trouver le but qu'en fonction de A, mais n'a rien retenu de B. b) Exemple d'occultation : l'apprentissage est réalisé en fonction des deux amers, mais un seul permet de mieux prédire l'emplacement du but (ici, A est plus proche de la plateforme). Lors du test avec B seul, l'animal n'ira pas plus vers l'emplacement de la plateforme que les autres quadrants car il n'aura appris à localiser le but qu'en fonction de A.

également qu'il y a une préférence graduelle croissante pour l'amer le plus proche du but.

Redhead et Hamilton (2007) démontrent également que, dans une piscine de Morris virtuelle dans laquelle est située une plateforme fixe, les humains accordent une valeur associative hiérarchiquement décroissante à une balise, à une configuration géométrique, puis à un amer distal (un mur coloré, cf. aussi Waller et Lippa, 2007).

Blocages et occultations entre amers de même type. Les phénomènes de blocages entre amers de même catégorie (entre amers proximaux ou entre amers distaux) sont curieusement les moins nombreux à avoir été étudiés. Stahlman et Blaisdell (2009) rapportent un cas de blocage d'un amer proximal par un autre (des formes colorées), dans un appareil constitué de coupelles où se trouve de la nourriture dans l'une d'entre elles. Ils ont facilement obtenu un blocage de l'un des amers, dans une procédure semblable à celle de la Figure 2.6 a.

Rodrigo et al (1997) et Biegler et Morris (1999) ont eu plus de difficulté à obtenir un blocage entre deux amers proximaux, ayant à réaliser un pré-entraînement en présence d'une configuration de plusieurs autres amers avant d'obtenir un blocage d'un nouvel amer, l'entraînement avec un seul n'ayant pas suffi.

Blocages et occultations entre amers proximaux et distaux. De nombreuses expériences montrent la capacité des indices proximaux à bloquer ou occulter les indices distaux. La première d'entre elle est celle de Chamizo et al (1983) qui montre un effet d'occultation entre un but visible et les amers distaux situés autour de la piscine de Morris. Redhead et al (1997) démontrent aussi que les amers distaux ne sont pas bloqués si les amers proximaux ne permettent pas de retrouver le but.

Roberts et Pearce (1999) précisent les conditions dans lesquelles ce blocage intervient : différents groupes doivent apprendre à rejoindre un but, mais celui-ci sera déplacé plus ou moins rapidement, et un amer différent pourra lui être substitué dans certains groupes. Les résultats montrent que le blocage n'intervient que lorsque les amers distaux sont disqualifiés dès le début de l'expérience (e.g., lorsque la plateforme est déplacée tous les essais) et que l'amer est le même du début à la fin de l'expérience. Ils en déduisent alors qu'un processus attentionnel dirige les animaux vers l'amer proximal, les conduisant à progressivement ignorer les amers distaux. Les mêmes conclusions sont retrouvées par Timberlake et al (2007) qui observent un blocage de l'apprentissage des amers distaux par l'amer proximal lorsque ce dernier est situé sur la plateforme, tandis qu'il disparaît lorsque l'amer proximal est situé à côté de la plateforme.

Pour Gibson et Shettleworth (2003), dans un labyrinthe "sec" (*dry maze*) et Gibson et Shettleworth (2005), dans un labyrinthe radial, le blocage des amers distaux serait plutôt dû à l'apprentissage progressif d'une "non pertinence" de ces amers (*learned irrelevance* lorsque d'autres prédisent correctement un but).

On retrouve également des cas réciproques où les amers distaux bloquent ou occultent les amers proximaux, lorsque ces derniers sont moins saillants : Diez-Chamizo et al (1985) montrent que des rats, dans un labyrinthe radial, discriminent très rapidement deux bras en fonction des seuls amers proximaux. Lorsqu'ils ont accès aux deux types d'amers, proximaux et distaux, ils occultent les amers proximaux, mais apprennent plus lentement la tâche.

Apprentissage d'une stratégie de lieu. Une stratégie de lieu suppose l'existence d'une représentation spatiale de l'environnement, ne serait-ce que pour la capacité des animaux à inférer de nouvelles trajectoires pour rejoindre le but. Son existence n'est pas exclue dans ce contexte, mais son apprentissage est supposé être dû à l'association, non pas d'un stimulus mais d'un groupe de stimuli avec une réponse menant à une récompense. Ce groupe serait ensuite traité comme un stimulus à part entière – et donc soumis aux effets "associatifs". Skinner et al (2003) montrent que l'atteinte d'un but invisible peut être reproduite par successions de stratégies simples apprises incrémentalement par compétition d'amers. Blaisdell (2009) et Pearce (2009) font l'hypothèse également que cette représentation est réalisée par un apprentissage associatif des amers distaux.

Whishaw (1991) conforte cette hypothèse en indiquant que les animaux ont de bonnes performances pour retrouver une plateforme cachée indiquée par les amers distaux lorsqu'ils suivent une trajectoire stéréotypée, mais qu'ils ont de mauvaises performances pour inférer de nouvelles trajectoires.

2.3.2 Théorie cognitive et apprentissage spatial

2.3.2.1 Expériences de navigation animale démontrant des effets non "associatifs"

De nombreuses expériences indiquent donc qu'indéniablement certains amers gagnent par compétition une saillance particulière par rapport à d'autres, dans un apprentissage spatial. Cependant on a vu que ces effets sont parfois difficiles à obtenir. Dans quelques expériences ils ont même été impossibles à mettre en évidence.

Absences de blocage et d'occultation. Brown et Bowman (2002) ne trouvent pas d'occultation entre amers distaux et/ou proximaux dans une expérience où les rats sont uniquement entraînés avec des amers proximaux (des plots marqués parmi d'autres plots non marqués dans un labyrinthe "sec") mais réussissent à retrouver la nourriture pendant des tests avec tous les plots non marqués. De même McGregor et al (2004a), lors d'une tentative de démonstration d'occultation par des amers distaux de deux amers proximaux déterminant un angle équilatéral avec la plateforme d'une piscine de Morris, n'ont pas réussi à empêcher les rats d'utiliser ces deux amers en l'absence des amers distaux. De nombreuses expériences démontrent également l'absence d'effets associatifs entre des amers discrets et la configuration géométrique du dispositif (pour une revue, Pearce, 2009).

Potentialisation entre amers. Pearce (2009) reconnaît que beaucoup de tentatives de son équipe à chercher des effets de blocage et d'occultation ont révélé plutôt des phénomènes de facilitation entre amers. Par exemple Pearce et al (2001), en étudiant les interactions entre un amer proximal et la forme de la piscine de Morris, ont démontré que l'apprentissage du but s'en trouve renforcé. L'équipe de Pearce a aussi prouvé par beaucoup de protocoles que la couleur des murs, avec la configuration géométrique, pouvaient faciliter l'apprentissage de la position de la plateforme dans une piscine de Morris (revue dans Pearce, 2009).

2.3.2.2 Une intégration d'amers

Des expériences historiques. Même si ces expériences ont été critiquées par la suite, Blodgett (1929) et Tolman et Honzik (1930) ont émis l'hypothèse que les rats peuvent choisir *sans essais et erreurs* les trajets les plus performants dans un labyrinthe et qu'ils peuvent apprendre *sans récompense particulière* (par apprentissage "latent") comment s'orienter dans un environnement. Lorsqu'un but y est disposé, ils peuvent en effet le rejoindre aussi rapidement, et sans essais préalables, que des animaux qui auraient pu pratiquer l'environnement avec récompense. Les caractéristiques de cet apprentissage sont donc à l'opposé de celles d'un apprentissage associatif.

Morris et al (1982) montrent de même la capacité des rats à retrouver rapidement dans une piscine circulaire la plateforme cachée, sur la seule présence d'amers distaux.

Ces expériences historiques ont ouvert la voie à l'hypothèse d'une représentation interne de l'environnement, supposée être acquise non incrémentalement, par apprentissage latent et être utilisée comme une carte pour choisir "mentalement" un trajet à exécuter pour rejoindre un but (e.g., Tolman, 1948).

L'apprentissage spatial ne se ferait pas dans ce cas par associations stimulus conditionné-réponse-récompense, mais par associations entre stimuli (S-S), i.e., tous les indices environnementaux. En d'autres termes, prenant l'exemple d'une piscine de Morris, l'animal n'aurait pas appris quel trajet il doit faire entre un certain amer et la plateforme, mais les caractéristiques des lieux, compte tenu de la configuration de stimuli environnante. Lorsqu'il positionnera pour la première fois la plateforme, le but s'intégrera à sa représentation – il n'aura donc pas besoin d'entraînement par essais et erreurs – et il pourra la rejoindre de façon flexible par divers trajets. Cet apprentissage ne relève donc pas d'une "procédure", comme l'apprentissage associatif, il est dit "déclaratif" : l'animal a appris "quoi faire" et non "comment faire". Tolman et Honzik (1930) supposaient que des contiguïtés spatiales de stimuli seraient à l'origine de cette association : lorsque l'animal perçoit un lieu, en se déplaçant il "s'attend" à percevoir un lieu voisin. De proche en proche, de vastes configurations topologiques se construisent pour représenter l'environnement.

Les caractéristiques d'un apprentissage spatial déclaratif.

Ce type d'apprentissage est dit *flexible*, parce qu'il est rapide. Quelques minutes suffisent pour qu'un rat ait une représentation spatiale relativement fiable de son environnement (Morris, 1981). Cela a été démontré expérimentalement par Steele et Morris (1999), qui ont mis en évidence le fait que les rats rejoignent, après un seul essai exploratoire, un but caché déplacé régulièrement tous les 4 essais. La représentation ayant été apprise indépendamment d'une récompense, ils n'ont donc pas besoin de réapprendre plus longuement un nouveau trajet vers le but.

Son cadre de référence est *allocentrique*, car l'animal peut rejoindre un but caché de n'importe quel point de départ.

Il est dit "*cognitif*", parce que l'animal semble pouvoir choisir entre plusieurs réponses possibles avant d'effectuer réellement le trajet vers le but. C'est ce qui correspond à la capacité de *pla-*

nification. Elle a été avancée par Tolman (1948) chez les rats, dans une expérience classique de détours. Dans un labyrinthe qui contient trois chemins possibles vers le but, chacun de longueur différente, des obstacles sont ajoutés, bloquant successivement les chemins les plus courts. Dès qu'un tel obstacle est rencontré, les rats entraînés choisissent rapidement le chemin le plus court restant. Cette capacité de détour, démontrée chez d'autres mammifères (Thinus-Blanc, 1996), se restreindrait chez les rats aux zones de l'environnement déjà explorées.

Il serait "*coûteux*". En effet, Whishaw (1998) teste des rats dans un labyrinthe circulaire où se trouvent disposées 8 coupelles le long du mur et une ouverture vers leur nid. Dans chaque coupelle, une lourde boulette est disposée de façon à ce que les rats doivent la rapporter au nid. L'auteur démontre que certains rats utilisent une stratégie procédurale (au sortir du nid, ils parcourent circulairement et systématiquement les coupelles pour trouver la coupelle pleine) et d'autres vont tout droit vers une coupelle non encore explorée. Cette dernière stratégie est plus performante en terme de temps plus court passé hors du nid, mais est moins rapide que l'autre en terme de vitesse de locomotion. Les auteurs remarquent en effet que si le premier groupe va "au galop" chercher de la nourriture, les animaux du second s'arrêtent fréquemment "en inspectant l'environnement de leur tête", comme pour vérifier mentalement leur trajet. Ce temps de "réflexion" serait donc un coût attribué à une stratégie qui semble malgré tout plus optimale que la précédente, par exemple dans une situation de prédation.

Un appui par des bases neurobiologiques. L'existence chez les animaux d'une telle représentation mentale dite "carte cognitive" a suscité et suscite encore des controverses. Elle a été néanmoins confortée par la découverte de neurones sensibles à la position du rat dans son environnement, situés dans la région hippocampique (voir section 2.4) : les *cellules de lieu* hippocampiques, dont les activations sont corrélées à l'emplacement allocentrique de l'animal dans son environnement (O'Keefe et Nadel, 1978), seraient construites par intégration des amers environnants et avec des informations idiothétiques (Sharp et al, 1995; Whishaw, 1998). Les amers seraient donc intégrés de manière non compétitive : si un amer est ajouté à l'ensemble après un premier apprentissage, il sera intégré aussi à la carte et ne sera pas bloqué (O'Keefe et Nadel, 1978; Poucet et Save, 2009).

Plus récemment, des cellules de direction de la tête – qui donnent une notion d'orientation –, et de grille – qui donnent une notion de distance –, découvertes dans des zones connectées à l'hippocampe, ont complété les bases neurobiologiques de la navigation chez le rat (Ranck, 1984; Taube et al, 1990; Hafting et al, 2005; Wiener et Taube, 2005).

2.4 Un apprentissage spatial dépendant de systèmes neuraux distincts

Même si des expériences semblent conforter les paradigmes associatif ou cognitif, des partisans de ces deux conceptions ne sont cependant pas contre l'hypothèse de l'existence de structures nerveuses distinctes gérant spécifiquement certains amers, proximaux ou distaux, sous-tendues par des systèmes neuraux différents. Les structures les plus souvent mentionnées (hippocampe,

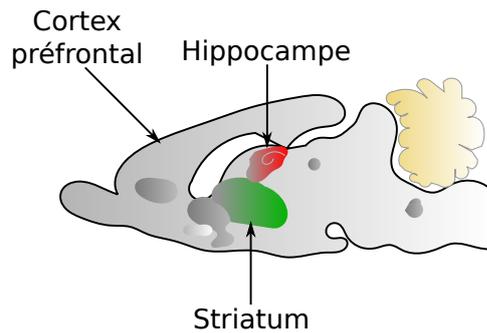


Figure 2.7 – Schéma des structures neurales de l'encéphale du rat impliquées dans la navigation.

cortex préfrontal, striatum) sont indiquées sur la Figure 2.7.

2.4.1 Des changements de stratégies.

Changements ponctuels. L'utilisation séquentielle où le changement brusque de stratégies induisent le fait que la pertinence de certains amers peut changer brusquement selon les circonstances, ce qui va à l'encontre d'un ré-apprentissage incrémental de nouveaux amers. Des expériences ont par exemple montré que les rats étaient capables de changer de stratégies selon le contexte courant. Ainsi Hamilton et al (2004) mettent en évidence la présence d'une séquence entre stratégies de lieu, puis de guidage, pour rejoindre un but visible dans une piscine de Morris : les animaux ne le repèrent pas tout de suite et se baseraient sur les amers distaux pour en déduire une direction générale puis, une fois le but identifié, passeraient à la stratégie de guidage, ce que les auteurs détectent par le changement brusque de trajectoire opéré par les animaux. Clark et al (2007) retrouvent des résultats similaires : quatre amers sont suspendus au-dessus des différents quadrants de la piscine, la plateforme cachée se trouvant sous l'un d'eux. Les animaux apprennent à retrouver cette plateforme en choisissant d'abord le quadrant correct sur la base des indices distaux puis retrouvent l'emplacement exact de la plateforme en allant sous l'amer de ce quadrant.

Tamara et al (2010) comparent les distances des points de départ de rats par rapport à la plateforme d'une piscine de Morris : le comportement des animaux y est analysé en terme de navigation égocentrique et allocentrique. Ils montrent que plus les rats sont positionnés près de la plateforme, plus ils utilisent une stratégie égocentrique. Plus elle est éloignée, plus ils utilisent une stratégie allocentrique. De même, des rats sont capables d'atteindre une plateforme visible, puis invisible lors de certaines sessions intercalées aux premières, avant de pouvoir rejoindre en un seul essai la plateforme visible, mais déplacée à un autre endroit (Devan et White, 1999). Ils auraient donc enchaîné des stratégies de guidage et de lieu, puis à nouveau une stratégie de guidage pour atteindre la plateforme déplacée en un essai.

Changements progressifs. Les changements progressifs de stratégies, non pas pendant un essai mais pendant une expérience, ont aussi été abondamment étudiés. Par exemple, il a été

montré que, dans une tâche qui consiste à rechercher une plateforme indicée mais déplacée tous les 4 essais, des rats passent progressivement d'une stratégie de lieu à une stratégie de guidage (Pearce et al, 1998). Un tel changement progressif d'une stratégie cognitive vers une stratégie procédurale (*goal-directed* versus *habit*) a fait l'objet de très nombreux articles en neurosciences (voir Khamassi, 2007, pour une revue). Ce changement est souvent associé avec les effets du surapprentissage (*overtraining*) : à force de répéter la même trajectoire, les rats montrent une baisse de la flexibilité de leur comportement (e.g., insensibilité à la dévaluation du but). Les facteurs influençant ce changement sont divers, mais la stabilité de l'environnement semble favoriser l'émergence d'une stratégie procédurale (e.g., Atallah et al, 2006).

Les expériences réalisées sur des animaux intacts ne peuvent pas se prononcer sur la correspondance entre systèmes neuraux et stratégies. Celles impliquant des inactivations de structures nerveuses spécifiques peuvent mettre en évidence le contrôle de stratégies par des systèmes distincts. Là encore, nous ne rappellerons que des points essentiels.

2.4.2 Des fonctions différentes entre systèmes neuraux.

De nombreuses expériences d'inactivation montrent notamment que le rat lésé sélectivement ne perd qu'une partie de ses capacités navigationnelles. Elles ont mis en évidence le support de l'une ou l'autre des stratégies par différents systèmes, impliquant la gestion par ceux-ci de modalités distinctes (e.g., cadre de référence, type d'amers, etc.). La majorité de ces expériences est fondée sur la dichotomie entre les stratégies indépendantes de l'hippocampe et les stratégies qui en sont dépendantes.

Système de guidage et système de lieu. Les stratégies procédurales comme les stratégies de guidage se sont avérées inutilisables lors de lésions du striatum dorsolatéral (DLS, Packard et McGaugh, 1992; McDonald et White, 1994). Le cortex pariétal s'avère aussi essentiel au bon fonctionnement de ces stratégies procédurales (Kolb et Walkey, 1987), du fait de son rôle supposé dans le cadre de référence égocentré (Save et Moghaddam, 1996). Les indices visuels semblent d'ailleurs privilégiés, car une lésion du colliculus supérieur empêche les rats de s'orienter vers un amer (Goodale et al, 1978).

L'hippocampe et d'autres structures neurales avec lesquelles il a des connexions, tel le striatum dorsomédial (DMS) ou le cortex médial préfrontal (mPFC), seraient impliqués dans l'apprentissage de stratégies cognitives et des stratégies de lieu (Granon et Poucet, 1995; Yin et Knowlton, 2004; Yin et al, 2005; Grahm et al, 2008; Mizumori, 2008; Jankowski et al, 2009; White, 2009).

Entrées sensorielles. Les entrées sensorielles gérées par ces systèmes seraient spécifiques. Ainsi une lésion du cortex pariétal empêche les rats d'utiliser les indices proximaux (Save, 2000), tout comme les lésions du DLS (Packard et McGaugh, 1996; Devan et White, 1999).

En revanche une lésion du cortex entorhinal empêche les rats d'utiliser les amers distaux, tout comme les lésions de l'hippocampe du fornix-fimbria et du striatum médial (Packard et McGaugh, 1996; Devan et White, 1999). Toutefois, des rats lésés de l'hippocampe peuvent apprendre à rejoindre une plateforme en présence des seuls amers distaux (Morris et al, 1982).

De même, un autre cas de lésion du cortex entorhinal n'a pas empêché les rats d'utiliser les amers distaux par une stratégie procédurale (Oswald et Good, 2000).

Il a été également démontré récemment que les cellules de lieu pouvaient être construites non seulement par intégration d'amers distaux, mais aussi en intégrant des amers proximaux, voire les deux types d'amers (Shapiro et al, 1997; Save, 2000; Renaudineau et al, 2007). L'hypothèse des entrées sensorielles spécifiques ne serait peut-être pas établie de façon certaine.

Cadre de référence. Le cadre de référence utilisé par ces systèmes semble bien délimité au premier abord : le DLS serait impliqué dans un référentiel égocentrique. De même, des lésions du cortex préfrontal médial sur des rats font disparaître l'expression d'une exécution égocentrique du trajet (Brasted et al, 1997; De Bruin et al, 1997; De Leonibus et al, 2005). L'hippocampe serait quant à lui impliqué dans un référentiel allocentrique (Oliveira et al, 1997).

D'autres expériences rendent un peu plus floues la dichotomie DLS/égocentrique et hippocampe/allocentrique. Des lésions de parties du DLS n'affecteraient pas forcément l'utilisation d'une stratégie égocentrique (Oliveira et al, 1997; Botreau et Gisquet-Verrier, 2010). L'hippocampe serait aussi nécessaire pour les séquences de navigation, même dans un référentiel égocentrique. Rondi-Reig et al (2006) ont montré en effet que des rats lésés de l'hippocampe ne pouvait pas apprendre une séquence d'associations procédurales simples (e.g., à la vue de tel amer, tourner à droite) au delà de 2 ou 3 associations successives, tandis que les rats contrôle apprenaient des séquences entières.

Caractéristiques de leurs apprentissages. Ces systèmes peuvent présenter des vitesses d'apprentissage différentes. Packard et McGaugh (1996) ont montré dans un T-maze entouré d'amers visuels que la stratégie de lieu, gérée par l'hippocampe, était acquise plus rapidement que la stratégie de guidage procédurale, indépendante de l'hippocampe. L'hippocampe et le cortex préfrontal seraient impliqués dans des apprentissages rapides (Squire et Zola, 1996; Eichenbaum et al, 2007).

Le DLS, impliqué dans un apprentissage de type associatif, serait lent et peu flexible : un changement de tâche rend caduque ces associations (Mishkin et al, 1984). La dopamine, un neuromodulateur, est d'ailleurs supposée jouer le rôle de signal d'erreur dans ce système, permettant de corriger incrémentalement les associations. Les expériences ont montré que le niveau de dopamine était dépendant de la récompense : lorsque celle-ci était inattendue, une forte augmentation de l'activité des neurones dopaminergique était observée juste après, baissant au fur et à mesure de l'entraînement. A l'inverse l'absence soudaine de récompense était suivie d'une baisse importante de cette activité (Schultz, 1998; Brown et al, 1999; Montague et al, 2004).

Même si dans la plupart des cas, les apprentissages impliquant le DLS sont relevés comme étant plus lents que ceux impliquant l'hippocampe, certaines tâches (comme la discrimination, i.e., le fait de distinguer deux stimuli différents, e.g., Broadbent et al, 2007) peuvent être gérées très rapidement par le DLS sans que l'hippocampe n'ait un rôle à jouer. Cette observation a été retrouvée en navigation : Martel et al (2007) ont montré que les stratégies de guidage (impliquant donc le DLS) pouvaient être apprises plus tôt que celles de lieu (impliquant l'hippocampe) si l'animal est préalablement familier avec le contexte. De même, le sur-apprentissage (*overtraining*) n'entraînerait pas forcément un changement d'une stratégie allocentrique vers une stratégie

égocentrique (Kealy et al, 2008).

2.5 Interactions entre systèmes neuraux

2.5.1 Systèmes neuraux apprenant en parallèle, mais pas indépendants

Apprentissages simultanés. Le corpus d'expériences précédentes a conforté l'hypothèse que les stratégies de navigation seraient gérées par des systèmes neuraux distincts, que ce soit au niveau des entrées sensorielles et de l'apprentissage (White et McDonald, 2002; McDonald et al, 2004a,b; White, 2004). L'étude de l'évolution du taux d'acétylcholine (ACh, un indicateur de l'intensité avec laquelle un système apprend) pendant l'acquisition d'une tâche a montré une forte corrélation entre les fluctuations de ces taux dans l'hippocampe et le striatum et la nature du comportement (Chang et Gold, 2003; McIntyre et al, 2003). Ainsi, l'étude des taux d'ACh dans les différents systèmes peut permettre d'évaluer leur contribution à l'apprentissage d'une tâche. Pych et al (2005) ont notamment remarqué que les taux d'ACh restaient élevés dans l'hippocampe alors même que l'animal effectuait un comportement S-R, suggérant que l'hippocampe participait toujours, d'une manière ou d'une autre, à la tâche. De plus, cette "activation" de l'hippocampe n'était observée qu'en présence de nombreux amers distaux mais non en leur absence. Ces résultats sous-entendent que l'hippocampe continuait l'apprentissage de la tâche alors même qu'il n'était pas impliqué dans le contrôle du comportement.

2.5.2 Sélection entre systèmes

La notion de systèmes neuraux apprenant des stratégies en parallèle implique une sélection quant à celle qui prendra réellement le contrôle sur les effecteurs – suivant le principe de la "voie comportementale finale commune" (*behavioral final common path*, McFarland et Sibly, 1975). D'une manière générale, le mPFC est impliqué dans la résolution de conflits : sa lésion empêche l'inhibition des mauvaises réponses dans une tâche et il échoue à proposer une nouvelle stratégie basée sur les indices contextuels actuels (Miller, 2001; de Wit et al, 2006).

Les notions de sélection *intra-modale* et *inter-modale* ont été désignées pour distinguer le changement de stratégies utilisées dans le même type de tâche ou dans des tâches différentes (Birrell et Brown, 2000). Dans le domaine de la navigation, ces modalités différentes ont été appliquées le plus souvent aux stratégies de guidage et aux stratégies de lieu. Une sélection intra-modale se fera donc entre stratégies de guidage ou entre stratégies de lieu, une sélection inter-modale se fera entre stratégies de guidage et de lieu.

Sélection intra-modale. Une sélection intra-modale peut être déclenchée par une tâche du type *reversal-learning*, qui correspond à toute situation d'apprentissage essais et erreurs dans laquelle un entraînement différentiel pour deux stimuli, respectivement récompensé-non récompensé (ou récompensé-puni), est inversé. Ragozzino et Choi (2004) effectuent des tests de reversal-learning entre des stratégies de guidage et entre des stratégies de lieu : des rats contrôles ou dont le DMS a été inactivé apprennent à rejoindre un but dans un labyrinthe en croix, but toujours indiqué selon les amers distaux. Ils étudient ensuite les performances de ces mêmes rats à rejoindre le but lorsque celui-ci a été déplacé (il n'est plus identifiable par les mêmes amers

distaux que lors de l'entraînement). Les rats "inactivés" se sont révélés moins efficaces pour apprendre le nouvel emplacement du but. En revanche, le DMS ainsi que le cortex orbito-frontal (OFC), interviendraient dans des tâches intra-modales, toutefois les lésions conjointes de l'OFC et de l'amygdale annuleraient les effets de la lésion de l'OFC seul sur une tâche de ce type (Kim et Ragozzino, 2005; Boulougouris et al, 2007). Selon Stalnaker et al (2007), ce résultat signifie que l'OFC ne participe donc pas directement à l'apprentissage, par exemple, en apprenant rapidement de nouvelles associations mais "prévient" les systèmes efférents pour que ceux ci corrigent leur représentation.

Sélection inter-modale. Les systèmes impliqués dans une sélection inter-modale seraient bien distincts de ceux engagés dans la sélection intra-modale. Ragozzino et al (1999) montrent que la lésion des aires prélimbiques (PL) et infralimbiques (IL) du cortex préfrontal empêche les changements de stratégies de lieu vers stratégies de guidage (voir aussi Granon et Poucet, 1995). Killcross et Coutureau (2003) précisent que ces deux aires contrôlent chacune l'expression d'un apprentissage déclaratif pour le PL, procédural pour l'IL. Rich et Shapiro (2009) se sont intéressés aux rôles plus spécifiques de ces aires. Le PL anticiperait les performances de l'apprentissage et aiderait donc à préparer le changement de stratégie, tandis que l'IL semblerait aider à l'établissement de la nouvelle stratégie. Le mPFC serait donc nécessaire pour coordonner les différents systèmes d'apprentissage.

Les sélections intra- et inter-modales seraient donc gérées par des systèmes distincts de ceux qui apprennent les stratégies. Les deux types de sélections seraient également gérés par des structures nerveuses distinctes.

2.5.3 Nature compétitive et coopérative des interactions

Il a été démontré que les systèmes neuraux sous-tendant des stratégies différentes peuvent avoir des interactions de nature compétitive (un pas vers la conciliation entre théories associative et cognitive), mais qu'ils peuvent aussi coopérer (e.g., Kim et Baxter, 2001; White et McDonald, 2002; Poldrack et Packard, 2003; Gold, 2004; Hartley et Burgess, 2005).

La compétition est définie par le fait que la lésion d'une structure impliquée dans l'un des systèmes entraîne une amélioration de l'apprentissage de l'autre par rapport à la performance globale (Figure 2.8 a). La coopération est définie par une dégradation des performances de l'apprentissage global, suite à la lésion de l'un des systèmes (Figure 2.8 b).

Quelques exemples de compétition. Des expériences où des rats peuvent simultanément apprendre les deux types de stratégies, de guidage et de lieu, mettent en évidence le fait que des animaux lésés d'une certaine structure neurale sont plus efficaces que les animaux intacts. Ainsi McDonald et White (1994) et Devan et White (1999) entraînent les rats à retrouver une plateforme visible et stable avec accès aux amers distaux dans une piscine de Morris, puis mettent ensuite les stratégies en compétition en déplaçant la plateforme visible. Les résultats montrent que les rats intacts ont une performance moins grande que des rats hippocampiques. Une stratégie de lieu pourrait donc être en compétition avec une stratégie taxon (voir également Lee et al, 2008, pour une observation semblable de compétition entre les deux stratégies).

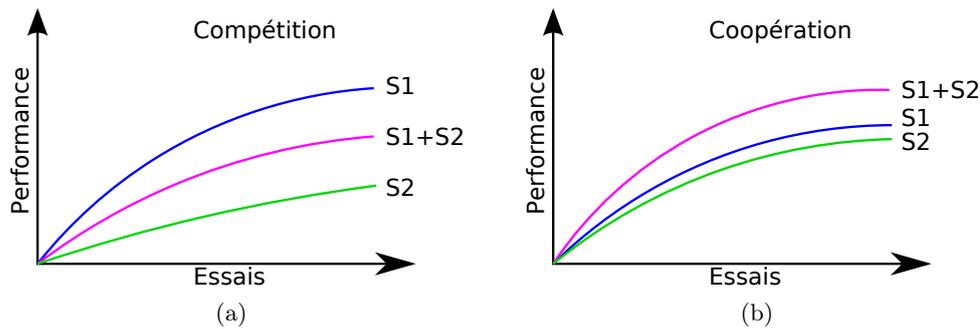


Figure 2.8 – Exemples schématiques de compétition (a) et de coopération (b) entre deux systèmes S1 et S2. Dans cette interaction compétitive, la lésion du système S2 entraîne une amélioration des performances du système S1 par rapport aux performances de S1+S2. Dans l'interaction coopérative, la lésion du système S1 ou S2 réduit leurs performances par rapport aux performances S1+S2. D'après Kim et Baxter (2001)

Pearce et al (1998) déclenchent ce conflit dès le début de l'expérience en déplaçant la plateforme régulièrement, après un petit nombre d'essais, de sorte à ce que seul l'amer proximal puisse prédire l'emplacement du but. Les rats contrôles utilisent d'abord une stratégie de lieu puis passent progressivement vers une stratégie de guidage, tandis que les rats hippocampiques utilisent immédiatement une stratégie de guidage. Une stratégie de guidage aurait ici été en compétition graduelle et décroissante avec la stratégie de lieu.

Par ailleurs, des injections de lidocaïne dans l'hippocampe, inactivant cette structure nerveuse, ont empêché des rats de développer une stratégie de lieu, mais ont facilité l'apprentissage d'une stratégie de guidage dans un labyrinthe en croix (Chang et Gold, 2003).

Quelques exemples de coopération. Nous pouvons ranger dans cette catégorie les expériences de l'équipe de Pearce (2009), mentionnées précédemment dans ce chapitre, montrant des potentiations entre amers dues à des coopérations entre stratégies différentes.

Un autre exemple de coopération plus subtil est celui de Micheau et al (2004) qui, après avoir entraîné intensivement des rats dans une piscine de Morris, les obligent à nager pendant un moment autour de la plateforme avant d'y grimper (*dwelling response*). Une stratégie de lieu est censée permettre d'aller vers la plateforme, une stratégie procédurale (indépendante de l'hippocampe) est censée régir la nage forcée autour de la plateforme. D'après les auteurs, le système hippocampique instruirait le système non hippocampique de "quand" et "où" il faut commencer la stratégie procédurale : si le rat a une incertitude quant à la position de la plateforme, il ne déclenche en effet pas sa stratégie *dwelling*.

Coopération et compétition. Pych et al (2005) démontrent, dans une expérience déjà citée plus haut, qu'il peut y avoir en même temps compétition et coopération entre systèmes. Des rats doivent apprendre soit une stratégie de guidage, soit une stratégie de lieu. Lorsque l'environnement est riche en amers distaux, le système hippocampique semble coopérer avec le système striatal pour l'apprentissage de la tâche procédurale : ils ont tous les deux un niveau

élevé d'ACh. Lorsqu'il y a peu d'amers distaux, les systèmes entreraient en compétition : le système hippocampique, peu performant, ralentirait l'apprentissage du système striatal.

Indépendance. D'après Cheng (1986) et Gallistel (1990), un module neural extrairait des informations géométriques de la vision de l'environnement, en excluant les autres informations (telles que couleur, luminosité, etc.). Cheng (1986) a mis en évidence une absence d'interactions entre la stratégie contrôlée par ce module, notamment gérée par l'hippocampe (McGregor et al, 2004b), et les stratégies procédurales basée sur la compétition d'amers (Cheng, 1986; Hayward et al, 2003; Redhead et Hamilton, 2007; Batty et al, 2009), même lorsqu'un amer (une balise) est sensiblement plus saillant que la géométrie (e.g., Redhead et Hamilton, 2007).

Un même constat est effectué pour les informations provenant des parois du dispositif ou limites d'une scène (Doeller et al, 2008; Doeller et Burgess, 2008).

2.6 Conclusion du chapitre

Nous avons présenté dans ce chapitre les notions essentielles, évidemment non exhaustives, qui caractérisent l'apprentissage spatial. Les informations environnementales nécessaires en entrée ainsi que les deux principales conceptions sur ses mécanismes – dépendant ou non de règles qui sous-tendent aussi tout conditionnement opérant – ont été résumées. L'une de ces conceptions suppose une compétition entre tous les amers pour le contrôle d'un comportement permettant d'avoir une récompense, l'autre suppose une intégration des amers environnementaux dans une représentation cohérente de l'environnement, acquise indépendamment de la récompense. Un corpus d'expériences conforte l'un ou l'autre de ces points de vue.

Ces deux paradigmes peuvent être cependant conciliés par l'hypothèse de systèmes neuraux séparés, apprenant différemment et en parallèle, pouvant être ancrés chacun sur des informations environnementales différentes (amers plus ou moins proches du but), ayant un cadre de référence différent aussi (égocentrique ou allocentrique). Ces systèmes pourraient être en interaction soit compétitive, ce qui conforte la première conception, soit coopérative, ce qui conforte la seconde – respectivement diminuant ou augmentant les performances globales lorsqu'ils sont tous deux en activité – ou indépendants.

Ils pourraient s'exprimer successivement, par sélection de l'un ou de l'autre, lorsque les circonstances l'exigent, comme le montre l'influence de l'environnement sur le comportement des animaux. Des structures spécifiques, non redondantes avec celles impliquées dans leurs apprentissage, comme semble l'indiquer les expériences de sélection intra- et extra-modale, les sélectionneraient en fonction de leur efficacité dans un contexte pertinent.

Cependant beaucoup de ces points ne sont pas établis expressément, comme l'ancrage spécifique sur certains amers, les cadres de référence ou la nature exacte de l'apprentissage régissant la représentation spatiale intégrant les amers, puisque des effets issus de règles associatives y sont avérés.

Le fait de pouvoir contrôler certains des facteurs précédents et d'avoir accès à des variables qui seraient cachées dans une analyse comportementale accorde aux modèles computationnels un avantage pour contribuer à considérer différemment les deux théories qui s'opposent. Dans le prochain chapitre, nous passerons en revue deux catégories de modèles, dont les auteurs respectifs n'ont pas encore eu beaucoup d'échanges réciproques. Nous les avons classés en modèles d'interactions entre amers, émergeant plutôt à la théorie associative, et en modèles de systèmes parallèles, dans lesquels des modules séparés prennent en compte les deux types d'apprentissages. Nous avons restreint ces derniers aux modèles de navigation bio-inspirés. Cette revue permettra de dégager les avantages et limites de chacun pour comprendre mieux divers effets de l'apprentissage spatial, afin de proposer un nouveau modèle qui pourrait mieux contribuer à l'objectif souhaité.

Chapitre 3

Modèles computationnels d'interactions entre amers et entre systèmes de navigation

Dans une première section nous présentons les principes et modèles issus de la théorie associative expliquant les compétitions entre amers. Ensuite, les principes et modèles appliqués aux différentes stratégies de navigation seules sont passés en revue, avant de présenter les modèles dévolus à leur sélection.

3.1 Modèles d'interactions entre amers

Les modèles computationnels présentés ici sont tous issus de la théorie associative, et de fait, ne font pas de distinction entre différents systèmes sous-tendant différentes stratégies. Ils font l'hypothèse que ce sont les amers présents dans l'environnement qui entrent en compétition pour le contrôle du comportement de l'agent. La question qui se pose alors concerne les modalités de cette compétition, i.e., les conditions qui favorisent un amer plutôt qu'un autre.

Certains fondent leurs algorithmes sur la règle de l'apprentissage associatif de Rescorla et Wagner (1972), d'autres sur le calcul bayésien de l'incertitude liée à la prédiction de récompense.

Nous présentons ici les modèles utiles dans leur application à la navigation. Nous ne rentrons cependant pas dans une analyse détaillée car, au-delà des améliorations successives ayant permis à certains de ces modèles de reproduire un nombre impressionnant de phénomènes comportementaux, ce sont avant tout des modèles analytiques qui, en tant que tels, n'ont pas pour objectif de fournir des prédictions concernant les fonctions de structures nerveuses spécifiques. Nous entendons par là que ces modèles dérivent d'une conception presque exclusivement behavioriste dont la vision des systèmes neuraux engagés peut se résumer à la boîte noire de Skinner (1938).

3.1.1 Règle de Rescorla-Wagner (Rescorla et Wagner, 1972)

La règle de Rescorla-Wagner a été utilisée à l'origine pour modéliser le conditionnement classique (Stimulus-Réponse), c'est à dire l'association d'un stimulus conditionné CS à un stimulus non conditionné US. Toutefois, les appliquer au conditionnement opérant (association d'un CS avec une réponse) ne relève que de l'échange des variables associées à l'US avec celles associées à une réponse. L'idée sous-tendant cette règle est qu'un stimulus non conditionné a une valeur d'association V à un stimulus conditionné qui est limitée par un paramètre λ . La mise à jour est réalisée à l'essai $n + 1$ entre un CS A et un US B :

$$\Delta V_A = \alpha_A \beta_B (\lambda_B - V_{\text{total}}) \quad (3.1)$$

où α_A et β_B sont des taux d'apprentissage dépendant respectivement du CS A et de l'US B, λ_B est la valeur d'association maximale que peut obtenir l'US B et V_{total} représente la somme des associations entre tous les CS présents et l'US B.

Ainsi le phénomène de blocage (*blocking*) est-il expliqué par le fait que le premier stimulus conditionné A a utilisé toute l'association disponible du stimulus non conditionné :

$$V_{\text{total}} = V_A = \lambda_B \quad (3.2)$$

et donc le second stimulus conditionné CS C ne peut plus y être associé, car la mise à jour sera pondérée par :

$$\lambda_B - V_{\text{total}} = 0 \quad (3.3)$$

De même l'occultation (*overshadowing*) d'un amer A sur un amer B, sera du à un α_A de valeur plus importante que α_B , et V_A atteindra donc plus vite la valeur critique. On pourra se reporter à Miller et al (1995) pour une revue des comportements reproduits par la règle de Rescorla-Wagner.

Cette règle connaît un grand succès par sa capacité à expliquer de nombreux phénomènes de psychologie comportementale de par sa simplicité. Elle fut naturellement appliquée à l'apprentissage spatial (associations entre amers environnementaux et actions pour rejoindre le but) pour expliquer ces mêmes phénomènes, suivant un certain nombre d'adaptations pour expliquer un nombre de plus en plus important d'effets associatifs (inhibition latente, facilitation...).

Toutefois, ce modèle montre certaines limites, comme la nécessité pour un amer d'être présent pour être réévalué, ou l'approche exclusivement dirigée vers le but de l'apprentissage (i.e., la confiance est attribuée à un amer uniquement en fonction des tentatives passées de l'animal à trouver le but en fonction de cet amer), ce qui a fait échouer le modèle à reproduire certains résultats expérimentaux (comme la *retrospective revaluation* que nous verrons par la suite, voir Pearce et Bouton, 2001, pour une revue des limites de ce modèle et de ses améliorations successives).

Les deux sous-sections suivantes présentent deux pistes récemment explorées pour améliorer les performances du modèle. D'une part, la question du niveau de la compétition des amers : l'hypothèse du comparateur, inspiré de la théorie de l'attention (Mackintosh, 1975) propose de situer cette compétition au niveau de la sélection et non de l'acquisition d'une stratégie. D'autre

part, l'évaluation des indices en fonction d'autres critères que leur seule capacité à prédire le but : le cadre bayésien permet une évaluation "a priori" des indices, au sens où il n'est pas nécessaire d'atteindre le but pour commencer leur évaluation.

3.1.2 L'hypothèse du comparateur (Miller et Matzel, 1988)

La règle de Rescorla-Wagner a été mise en défaut par un certain nombre de résultats comportementaux dont le point commun est la possibilité de modifier les forces associatives de certains indices sans que ceux-ci ne soient nécessairement présents. Par exemple Kaufman et Bolles (1981) montrent qu'après un conditionnement en présence de 3 stimuli (bruit, lumière et choc électrique), un apprentissage suivant en présence d'un seul stimulus (lumière) améliore l'association entre les deux stimuli restants (bruit et choc électrique). De fait ces phénomènes, nommés *retrospective reevaluation*, ne peuvent être simplement reproduits par la règle de Rescorla-Wagner puisque cette dernière n'explique pas la mise à jour de l'association entre deux stimuli alors même que ceux-ci sont absents.

L'hypothèse du comparateur Miller et Matzel (1988) propose alors que les associations entre deux amers sont indépendantes les unes des autres, i.e., que l'association entre un stimulus conditionné A et un stimulus B sera apprise indépendamment de l'association entre A et un autre stimulus C. Ainsi, en présence des trois stimuli, les différentes associations (A-B, A-C) seront comparées, et l'association de plus forte valeur sera sélectionnée. Si l'on reprend comme exemple l'expérience de Kaufman et Bolles (1981), l'amélioration de la réponse en présence des deux stimuli *bruit* et *choc électrique* sera alors fonction des règles suivantes : $\Delta V_{A,B}$ représente l'association entre les stimuli A et B, au sens du potentiel de A de rappeler la représentation de B.

$$\Delta V_{A,B} = AB(1.0 - V_{A,B}), \quad (3.4)$$

où $0 < A \leq 1$ et $0 < B \leq 1$.

Lorsque A est présenté sans B (et inversement), la mise à jour de $\Delta V_{A,B}$ est alors réalisée comme suit :

$$\Delta V_{A,B} = -kAV_{A,B}, \quad (3.5)$$

avec k un paramètre d'apprentissage.

La valeur d'une association entre un amer X et une récompense attendue O sera alors réalisée en soustrayant à l'association directe $V_{X,O}$ les associations intermédiaires entre les autres amers et X et les autres amers et O :

$$R_X = V_{X,O} - k_2 f\left(\sum_j V_{X,j} V_{j,O}\right), \quad (3.6)$$

où $\sum_j V_{X,j} V_{j,O}$ est le comparateur, k_2 déterminera la proportion maximum qui sera soustraite par le comparateur, et $f(x) = x$ si $x \leq 1$, $f(x) = 1$ sinon.

De fait, si l'amer de plus forte association vient à disparaître, alors l'amer restant qui aura l'association la plus forte sera choisi, contrairement à l'hypothèse associative pure, où toutes les autres associations auraient été dévaluées.

Ce modèle sera amélioré successivement jusqu'à prendre en compte des phénomènes complexes

comme les interactions de n^{eme} ordre (associations intermédiaire entre amers, Denniston et al, 2001; Stout et Miller, 2007). Notons toutefois que le point important de l'hypothèse du comparateur est l'idée que les acquisitions des associations ne peuvent pas être bloquées, comme dans la règle de Rescorla-Wagner, mais que les phénomènes de blocage proviennent d'une sélection effectuée entre ces associations : C'est l'association de plus forte valeur qui sera sélectionnée mais elle n'empêchera pas les autres associations d'être apprises.

3.1.3 Règle de sélection bayésienne

Si la règle de Rescorla-Wagner permet de lier la capacité d'un amer à prédire une récompense avec la force d'association que celui-ci recevra, d'autres facteurs concernant les amers ne sont pas pris en compte tels que leur taille et la stabilité de leur attribut (forme, couleurs, etc.). De fait, un élément sera d'autant plus utile pour repérer un but qu'il sera visible de n'importe quel lieu – ce qui implique qu'il soit suffisamment éloigné du sujet et de taille importante et "invariant" – que les attributs permettant son identification soient stables au cours du temps. On peut ainsi calculer la confiance que l'on injectera dans un amer en estimant sa variance, au sens bayésien du terme. La règle de Bayes permet justement de pondérer les amers selon leur variance : supposons que deux amers X et Y , respectivement de variances σ_X^2 et σ_Y^2 soient présents dans l'environnement. Les poids W_X et W_Y qui leur seront attribués seront alors :

$$W_X = \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}, \quad (3.7)$$

et

$$W_Y = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2}, \quad (3.8)$$

c'est à dire que la confiance que l'on attribuera à un amer sera inversement proportionnel à sa variance (Cheng et al, 2007). Il est aussi possible, lorsque deux amers sont proches, de les combiner. Ce qui explique pourquoi un certain nombre d'amers de petite taille peuvent être préférés à un amer de grande taille, qui aurait dû être logiquement préféré selon la règle de Rescorla-Wagner.

Ainsi Ratliff et Newcombe (2008) montrent des résultats contradictoires obtenus dans le cadre du module géométrique : dans certains cas, l'utilisation des informations géométriques n'est pas bloquée par des indices visuels (Cheng, 1986), alors qu'elle l'est dans d'autres (Horne et Pearce, 2009). Cela peut être expliqué par une pondération bayésienne des indices visuels et géométriques. De fait, la pondération bayésienne des indices est plus complémentaire que concurrente des règles associatives "radicales" présentées précédemment et peut permettre à ces dernières de reproduire alors les résultats initialement contradictoires, comme l'absence de blocage ou d'occultation dans certaines conditions (Miller et Shettleworth, 2007; Sturz et Bodily, 2010).

3.1.4 Avantages et limites des modèles d'interactions entre amers

La plupart des modèles dépendants de la théorie associative n'envisagent l'apprentissage spatial que par le truchement d'une compétition d'indices environnementaux. Tous les indices sont

censés être appris par la même règle procédurale. S'ils peuvent expliquer les cas de blocages ou d'occultations d'indices dans certaines conditions de navigation, ils ne pourraient pas expliquer le fait que l'intégration de plusieurs indices puisse améliorer la navigation, ou que différentes stratégies soient disponibles rapidement lorsque les conditions l'exigent. Nous avons vu dans l'état de l'art précédent un certain nombre d'expériences qui démontrent, d'une part, que des indices peuvent être appris différemment et, d'autre part, qu'ils peuvent être appris en parallèle de telle sorte que les stratégies puissent être utilisées séparément. Enfin, comme remarqué en début de section, ces modèles sont avant tout analytiques : ils proposent des règles mathématiques permettant de reproduire avec succès un certain nombre de comportements observés mais donnent peu d'indication sur leur implémentation possible par les systèmes neuraux engagés (même si, on le verra dans la partie suivante, certaines de ces règles peuvent être adaptées par des modèles plus bio-mimétiques).

Cependant, la proposition de pondérer de manière bayésienne les différents indices, couplée à sa capacité à prédire des phénomènes habituellement difficiles à reproduire (e.g., module géométrique) peut donner des indications quant au traitement des amers en amont de leur évaluation, lors de l'exécution de la tâche proprement dite.

3.2 Modèles de systèmes parallèles

Une grande partie des modèles computationnels bio-inspirés dévolus spécifiquement à la navigation ont comme hypothèse l'existence de plusieurs stratégies de navigation sous-tendues par des systèmes séparés et mis à jour par des apprentissages différents. Ces stratégies se fondent directement sur les entrées sensorielles ou élaborent un modèle du monde, i.e., une représentation spatiale. Cependant peu de modèles se focalisent sur la sélection et les interactions entre stratégies. Nous passons d'abord en revue les différents apprentissages mis en œuvre et les principales stratégies modélisées, avant de décrire les modèles de sélection.

3.2.1 Les différents apprentissages mis en œuvre

3.2.1.1 L'apprentissage par renforcement (apprentissage procédural)

Dans les modèles présentés dans cette section, des stratégies de guidage et de lieu peuvent être apprises au moyen d'un apprentissage par renforcement procédural, dont nous décrivons brièvement le principe.

L'apprentissage par renforcement fait à l'origine partie des techniques d'intelligence artificielle, et consiste en un apprentissage par *essais et erreurs* : à chaque action réalisée, l'agent reçoit une récompense numérique R_t qui représente la conséquence de cette action. Le but de l'agent est alors de maximiser l'obtention de ces récompenses en sélectionnant à chaque pas de temps l'action adéquate. A l'inverse de l'apprentissage supervisé, l'agent ne reçoit donc à aucun moment d'information sur l'action que celui-ci aurait dû effectuer, mais va estimer la valeur de chaque action selon l'état s courant en fonction du seul signal de récompense.

Différence temporelle. En pratique, cela nécessite d'apprendre une fonction $V(s)$ qui va encoder la valeur de la récompense espérée pour chaque état que l'agent va rencontrer et l'affiner en fonction de la récompense effectivement obtenue. Pour cela, l'agent va mettre à jour la valeur $V_t(s)$ d'un état en lui ajoutant la différence entre la récompense obtenue R_t et la récompense espérée $V_t(s)$, appelée différence temporelle (*temporal difference* TD)

$$V_{t+1}(s) \leftarrow V_t(s) + \alpha \delta_t, \quad (3.9)$$

avec α le taux d'apprentissage et :

$$\delta_t = V_t(s) - R_t + \gamma V_{t+1}(s), \quad (3.10)$$

Ainsi, si l'agent arrive dans état de faible valeur, mais qu'une importante récompense R est alors délivrée, l'erreur TD sera positive et l'agent rehaussera la valeur de l'état. A l'inverse, un état de forte valeur rencontré alors qu'aucune récompense n'est délivrée se verra dévalué par une erreur TD négative.

Le facteur γ gère l'affaiblissement de la récompense (*reward discount factor*). Il sert à fixer de combien l'agent va regarder de l'avant : par exemple, si l'agent n'obtient pas de récompense dans l'état s_t mais prédit ensuite qu'un état s_{t+1} fournit une grande récompense, l'erreur de différence temporelle δ_t sera quand même positive, ce qui signifie qu'à partir de l'état s_t , il est possible d'obtenir une grande récompense. Plus γ sera élevé et plus cette anticipation sera importante. Si l'agent connaît sa fonction de transition, i.e., l'état d'arrivée pour chaque action effectuée à partir d'un état de départ, il peut ensuite choisir à chaque pas de temps l'action qui lui permettra d'arriver dans l'état de plus forte valeur.

On a donc deux modules qui composent cet agent : un module nommé "Critique" (*Critic*) va apprendre la fonction valeur $V(s)$ pour estimer la valeur de chaque état. L'autre module, "Acteur" (*Actor*), va apprendre la politique de l'agent, c'est à dire à choisir l'action ou la séquence d'action qui va maximiser les récompenses obtenues, en s'aidant de l'estimation fournie par la fonction $V(s)$ (Houk et al, 1995).

Q-learning. La méthode précédente découple l'apprentissage de la fonction valeur de la politique de l'agent. Or, il est possible d'apprendre ces deux fonctions simultanément en utilisant l'algorithme *Q-learning* (Sutton et Barto, 1998) et en estimant la valeur des couples *état-action*. Cet algorithme consiste à apprendre une fonction $Q(s, a)$ qui va estimer la récompense espérée lorsque l'action a est effectuée dans l'état s .

La mise à jour de cette fonction s'effectue alors comme suit :

$$Q_{t+1}(s, a) \leftarrow Q_t(s, a) + \alpha \delta_t, \quad (3.11)$$

avec α le taux d'apprentissage et :

$$\delta_t = Q_t(s, a) - R_t + \gamma \max(Q_{t+1}(s_{t+1}, a_{t+1})), \quad (3.12)$$

avec $\max(Q_{t+1}(s_{t+1}, a_{t+1}))$ la valeur maximum des Q-valeurs dans l'état s_{t+1} et γ le facteur d'affaiblissement de la récompense, dont la fonction est similaire à celle expliquée pour la différence temporelle mais appliquée aux couples état-action.

Dans la plupart de ces algorithmes, il est souvent nécessaire de parfois explorer l'espace d'états, c'est à dire d'effectuer volontairement une action sub-optimale, au cas où cette action permettrait ultérieurement d'obtenir de meilleures performances. Les méthodes utilisées pour estimer la fréquence de cette exploration (la question du ratio exploitation/exploration est un thème à part entière du domaine de l'apprentissage par renforcement) consiste à prendre un taux fixe ou variable selon le temps passé sans recevoir de récompense (*ϵ -greedy*) ou utiliser une sélection probabiliste des actions, l'action optimale recevant la plus forte probabilité (*softmax*).

Relation avec le comportement animal. L'apprentissage par renforcement a au départ été très souvent utilisé pour la simulation des conditionnements classique et opérant (Sutton et Barto, 1998), il montre en cela un certain nombre de similitudes avec la règle de Rescorla-Wagner présentée dans la section précédente. Par ailleurs, le signal dopaminergique qui a été mesuré dans certaines parties des ganglions de la base (substance noire compacte ou aire ventrale tegmentale) pourrait représenter le signal d'erreur de différence temporelle (Schultz, 2002), ou préparer le système à un état d'apprentissage facilitant la modification des liaisons cortico-striatales (Redgrave et al, 1999). Enfin, l'anatomie et le fonctionnement des ganglions de la base se prêtent aisément à l'analogie avec les systèmes Acteur-Critique (Houk et al, 1995; Montague et al, 1996; Schultz et al, 1997; Khamassi et al, 2005).

Les caractéristiques de l'apprentissage par renforcement montrent cependant certaines limitations. Sa vitesse d'apprentissage, à l'origine assez lente et son manque de flexibilité. Il est en effet très long de "désapprendre" une politique lorsque des changements environnementaux rendent inopérants la politique précédemment apprise et d'en apprendre une nouvelle. Sa vitesse d'apprentissage a pu être grandement améliorée, par exemple par l'adaptation du Q-learning proposée par Strösslin et al (2005) (voir section 3.2.2.2).

Pour remédier au défaut de flexibilité, la construction d'un "modèle du monde" donne à l'agent la capacité de simuler sa politique avant de l'effectuer dans le monde réel. Cette technique repose le plus souvent sur la planification, qui peut être mis en œuvre à l'aide de la recherche de graphe.

3.2.1.2 Recherche de graphe (apprentissage déclaratif)

Nous avons décrit l'algorithme précédent sans préciser comment était estimé l'espace d'états-actions disponible pour l'animat. Là encore, il s'agit d'une problématique à part entière, mais dans le cas de la navigation qui nous intéresse, deux principales options existent : câbler directement les entrées sensorielles aux actions de l'animat – mais il est dans ce cas impossible d'apprendre la stratégie autrement que par l'exécution des actions possibles – ou construire une représentation de l'environnement, la plupart du temps sous forme de graphe (i.e., des nœuds, représentant les états, connectés entre eux s'il existe une action permettant de passer de l'un à l'autre). Avec ce graphe, il est alors possible de "simuler" l'exécution d'une stratégie pour en connaître l'issue sans avoir besoin de le faire dans le monde réel, autrement dit de "planifier".

Comme le nom l'indique, la recherche de graphe nécessite la construction préalable d'un graphe

(dont nous verrons une construction possible au chapitre suivant), dont les nœuds sont des états connectés entre eux lorsqu'une action existe pour les rejoindre. L'avantage d'une telle représentation tient en l'existence de ces liens qui permet une mise à jour de la politique très rapide, ainsi que de la possibilité de planifier un chemin pour rejoindre un but. Pour ce faire, il est évidemment possible d'utiliser les algorithmes étudiés depuis plus de 50 ans, comme l'algorithme de plus court chemin (Dijkstra, 1959). Des versions plus biomimétiques de cet algorithme ont également été proposées, comme le principe d'activation-diffusion (Burnod, 1991, nous verrons son fonctionnement dans le chapitre suivant). Enfin les algorithmes d'apprentissage par renforcement ont également été appliqués aux espaces d'état-action représentés sous la forme de graphe (*model-based* RL, comme Dyna, Sutton, 1991).

3.2.2 Modélisation computationnelle des stratégies

Bien que nous ayons vu qu'il existe plusieurs catégories de stratégies de navigation (section 2.2), nous n'allons exposer ici que les deux principales confrontées dans la grande majorité des expérimentations, à savoir les stratégies de guidage et de lieu.

3.2.2.1 Stratégies de guidage

Les stratégies de guidages n'impliquent pas l'utilisation de représentation interne et sont le plus souvent modélisées selon des associations stimulus-réponse.

Taxon. Lorsque le but est visible, la stratégie la plus simple pour le rejoindre consiste à d'abord s'orienter dans la direction du but puis d'avancer jusqu'à l'atteindre, en fonction d'un cadre de référence égocentré (Trullier et al, 1997). Des modèles simples, sans apprentissage, comme celui de Girard et al (2005), proposent que l'entrée de cette stratégie soit composée d'un vecteur de proximité, par exemple 36 neurones codant chacun pour 10° , dont l'activation dépend de la proximité avec le but dans la direction donnée. Il suffit ensuite de minimiser l'orientation avec la direction correspondant au neurone le plus activé et d'avancer jusqu'à atteindre le but.

Bien que ce modèle ne propose pas d'apprentissage, il est très simple de lui en adjoindre un, comme le fait Sheynikhovich et al (2009), en apprenant les associations entre entrées sensorielles et actions d'orientations au moyen d'une variante de Q-learning (que nous présenterons en détail dans le chapitre suivant), où encore Guazzelli et al (1998), qui supposent que l'animat à accès à deux types de perceptions : les *affordances* qui représentent les orientations que l'animat peut prendre (i.e., au point de jonction d'un labyrinthe en T, il s'agira des bras) et la perception de la nourriture. Chacune de ces perceptions est représentée par une gaussienne autour de son orientation, et un apprentissage par renforcement est effectué pour réévaluer cette gaussienne, selon la récompense obtenue après exécution de l'action associée.

Khamassi et al (2005) proposent un système *Acteur-Critique* pour modéliser cette stratégie. Les entrées sensorielles font l'objet d'un traitement préalable catégorisant l'environnement observée selon certaines propriétés en plusieurs variables de haut-niveau, telles que la vision d'une couleur particulière, son orientation et sa distance. De même les actions qui seront effectuées par

la partie Acteur sont des comportements stéréotypés tels que “aller vers le blanc”, etc. Enfin, suivant l’architecture des ganglions de la base, le modèle comporte plusieurs modules Acteur-Critique qui seront progressivement spécialisés selon un apprentissage non supervisé (cartes auto-organisatrices, *growing neural gas*, ...), sur la base des entrées sensorielles : chaque module sera ainsi responsable d’un ensemble de lieux partageant des caractéristiques proches (e.g., la quantité et répartition des différents amers visuels). Le modèle montre un apprentissage efficace, mais relativement lent, dans le cadre d’un labyrinthe en croix, où l’animat doit trouver un but balisé situé à l’extrémité d’un bras.

Guidance Le modèle computationnel de Cartwright et Collett (1983) propose, pour expliquer le comportement de *guidance* des abeilles, que celles-ci se souviennent de l’emplacement d’un but par l’image (*snapshot*) de l’environnement à son endroit. Ainsi pour retrouver le but, les abeilles cherchent à minimiser la vue courante de l’environnement avec le *snapshot* alors mémorisé, la différence entre les deux vues permettant de proposer une orientation (à l’instar des algorithmes de descente de gradient). Le cadre de référence de cette stratégie est aussi égocentré : si l’animal a appris l’emplacement du but en partant toujours d’un même endroit, il devra réapprendre cet emplacement s’il part d’un autre endroit.

Direction (*Heading vector*). La stratégie *Heading-vector* telle que définie par, e.g., Pearce et al (1998), est l’apprentissage d’une certaine direction et distance d’un but par rapport à un ou plusieurs amers. La reproduction de cette expérience par le modèle de Chavarriaga et al (2005) (voir aussi Dollé et al, 2008) intègre cette stratégie "Direction". Elle est modélisée par une association entre entrées sensorielles et actions, tout comme le fait Sheynikhovich et al (2009), à la différence que le cadre de référence est ici allocentrique, simulant l’ancrage de cette stratégie sur un amer proximal ou distal.

Kubie et Fenton (2009) proposent un modèle computationnel de *heading vector* qui est étendu au retour au nid. Les auteurs prennent appui sur l’existence des cellules de direction de la tête pour expliquer, notamment, les comportements de retour au nid observés chez certains animaux comme les rats, où ceux-ci, après avoir exploré leur environnement de manière pseudo-aléatoire, montrent la capacité de retourner vers leur point de départ en ligne droite comme s’ils étaient capables de sommation de vecteurs. Ce modèle propose justement une telle accumulation : des variables, telles que la distance parcourue et l’orientation cumulée, sont mémorisées et mises à jours durant l’exploration de l’animat, dans ce que les auteurs appellent des *Head-Direction Accumulators* (HDA). Les HDA sont des cellules réceptives chacune à une direction préférée et qui sont cumulées au fur et à mesure de l’exploration de l’animat. La cellule la plus activée correspond alors à la direction globale qu’à pris l’animat depuis le départ. Lorsque ce dernier veut retourner à son point de départ, il lui suffit alors de prendre la direction opposée à celle ayant la cellule du HDA la plus activée, l’accumulation de sa valeur se faisant alors en sens inverse, jusqu’à que cette dernière redevienne nulle.

Des modèles similaires ont également été proposés par Samsonovich et McNaughton (1997) et Mittelstaedt (2000). Le premier ne propose pas d’implémentation des cellules de directions de la tête tandis que le second nécessite la présence d’une carte de bas niveau.

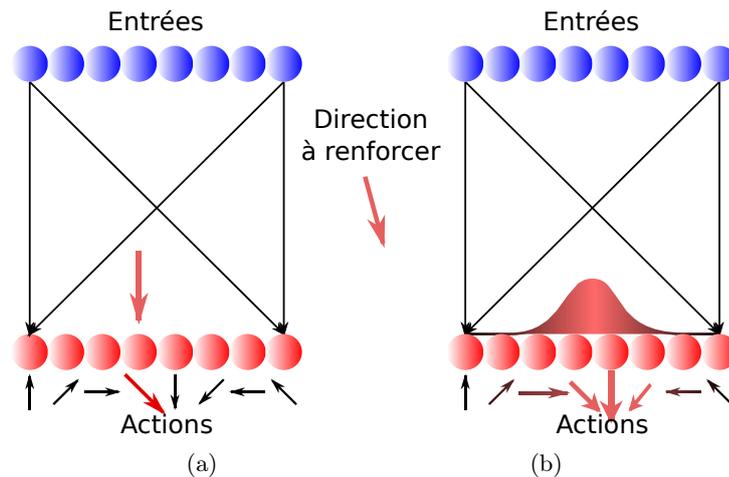


Figure 3.1 – (a) Règle de renforcement habituellement utilisée : on ne renforce que l'action choisie au temps t , L'action choisie au temps $t + 1$ sera l'action de plus forte Q-valeur. (b) Modification de Strösslin et al (2005) : le renforcement de l'action choisie (au centre) est modélisé par une gaussienne centrée autour de cette action. L'action qui sera choisie au temps $t + 1$ sera une interpolation des actions par leur Q-valeurs respectives.

3.2.2.2 Stratégies de lieu

Si l'on se réfère à la taxonomie proposée par Trullier et al (1997), deux sortes de stratégies de lieu peuvent être évoquées : la navigation Locale et la Planification. Nous décrivons brièvement le fonctionnement de ces deux stratégies et leur implémentations avant de discuter leur avantages et inconvénients. Tous les modèles implémentant les stratégies de lieu fonctionnent en mode allocentrique (entrées et actions d'orientation sont fonction d'un angle absolu indépendant de l'orientation de l'animat) et sont fondés sur une *intégration d'amers*.

Navigation Locale (*Place-Recognition-Triggerred-Response*) : la stratégie Locale associe à la reconnaissance d'un lieu une action locale d'orientation. Il n'est en effet pas supposé de connexion entre les différents lieux, il est donc impossible de prédire un lieu d'arrivée à partir d'un lieu de départ. Chaque orientation est calculée à chaque pas de temps. De fait, l'apprentissage nécessaire pour rejoindre un but est souvent long et inflexible : lorsque le but est déplacé, la trajectoire doit être entièrement réapprise.

La stratégie Locale est apprise au moyen d'un algorithme d'apprentissage par renforcement procédural. L'apprentissage s'effectue par la mise à jour des poids entre les entrées (le plus souvent, les activations d'une population de cellules de lieu) et les actions d'orientation. C'est le cas, notamment, du modèle d'Arleo et al (2001) : dans ce modèle, les cellules de lieux et d'action sont associées dans un réseau de neurones par des liens synaptiques, sur lesquels un apprentissage de type *Q-learning* est effectué. Les *Q-valeurs* des actions associées sont calculées comme la somme des entrées (i.e., l'activation courante des cellules de lieux) par les poids

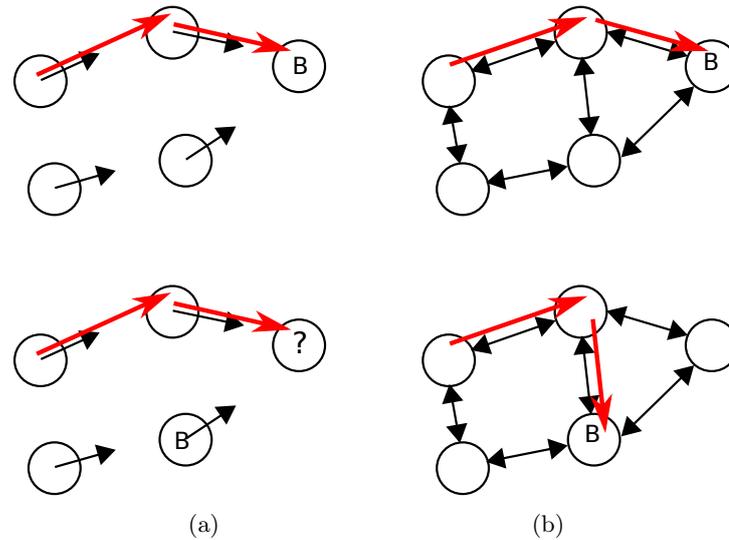


Figure 3.2 – (a) La stratégie Locale : chaque lieu associe une action d’orientation (flèche noire) unique sans information sur le lieu d’arrivée (B, le but), la trajectoire de l’animat (en rouge) persistera donc vers l’ancien emplacement, si le but est déplacé. (b) La stratégie Planification : l’information spatiale est représentée sous la forme d’un graphe topologique, les actions associées aux lieux (en noir) permettent de passer d’un lieu à un autre. Aussi, lorsque le but B est déplacé, la trajectoire pour le rejoindre est mise à jour dès que l’animat a pu le repositionner dans sa carte. D’après Trullier et al (1997)

synaptiques. Cet apprentissage est relativement rapide, puisque l’animat parvient à apprendre les trajectoires pour rejoindre un but donné en seulement 5 essais, soit la durée nécessaire généralement observée dans les expériences avec des animaux réels en piscine de Morris (Morris, 1981). Ce modèle souffre cependant d’un certain manque de généralisation, puisque la sélection de l’action est réalisée par un *winner-takes-all* (Figure 3.1 a). Les actions sont donc discrètes et la vitesse de l’apprentissage dépend directement du nombre d’actions possibles. De fait, plus le nombre d’actions possibles est élevé, plus la trajectoire apprise sera précise, mais au coût d’un apprentissage bien plus long. Strösslin et al (2005) remédie à ce problème en proposant un “profil de généralisation” : plutôt que de sélectionner l’action avec la plus forte *Q-valeur*, l’auteur propose de calculer l’action finale par une somme pondérée de toutes les actions par leur *Q-valeur* respective. De même l’apprentissage est non plus réalisé en fonction d’une action donnée, mais d’une orientation, le renforcement pour chaque action étant d’autant plus élevé que cette action propose une orientation proche de celle effectuée (Figure 3.1 b). Cette technique permet avec un nombre relativement limité d’actions (36 dans le cas présent) d’apprendre une trajectoire plus précise que précédemment, et ceci en moins de temps qu’il n’en n’aurait fallu avec un algorithme de *Q-learning* normal. Cette modélisation sera reprise dans Chavarriaga et al (2005).

Il reste toutefois que la stratégie Locale s'avère peu flexible, notamment dans le cas du *reversal learning* où la récompense (i.e., le but), qui a été reçue à un certain emplacement, doit être réapprise à un endroit où précédemment l'animal n'a pas reçu cette récompense (Figure 3.2, gauche). En effet, la stratégie Locale souffre du défaut commun aux techniques d'apprentissage par renforcement, qui tient à la difficulté d'apprendre une nouvelle politique après un premier apprentissage. En navigation, cette limitation est illustrée par l'incapacité de reproduire la courbe de performance de rats en simulation dans une piscine de Morris où le but est déplacé tous les 4 essais (Steele et Morris, 1999). Foster et al (2000) montrent en effet qu'un simple *TD-learning* est incapable de reproduire de tels résultats, la mémorisation des emplacements précédents de la plateforme perturbant l'apprentissage du nouvel emplacement.

Planification. Pour cette stratégie, une représentation allocentrée de l'environnement est construite à partir de connexions entre différents lieux, encodées sous forme de liens entre nœuds dans un graphe topologique.

La stratégie Planification repose en effet sur l'utilisation d'un graphe encodant des informations d'orientation nécessaire pour naviguer d'un lieu à l'autre (Fig. 3.2, droite). A l'inverse de la stratégie Locale, les actions sont apprises indépendamment du but sur un mode déclaratif, ce qui offre une plus grande flexibilité comportementale. Ainsi lorsque le but est localisé, une recherche de graphe permet facilement de planifier un chemin depuis le point de départ jusqu'à lui, ce qui évite l'apprentissage relativement lent observé sur les modèles de navigation Locale. En revanche, la densité du graphe et de ses connexions doit être maîtrisée pour éviter l'explosion combinatoire : si le graphe est complet, i.e., tous les nœuds sont connectés entre eux, le coût de la planification est exponentiel suivant le nombre de nœuds. De fait, dans la plupart des modèles, seuls les nœuds voisins sont reliés entre eux.

Les modèles de planification sont nombreux, et proposent des solutions très différentes pour les problèmes de création de cartes et de planification (voir Filliat et Meyer, 2003, pour une revue des ces modèles). Nous nous restreignons ici volontairement aux modèles bio-inspirés en séparant l'étape de création de la carte à celle de la planification proprement dite.

Création de la carte. Les modèles de Filliat et Meyer (2002), Gaussier et al (2002), Quoy et al (2002), Barrera et Weitzenfeld (2008) et Martinet et al (2008) apprennent une carte de l'environnement sur la base d'une simulation des cellules de lieu hippocampiques et font une différence entre la carte hippocampique et la carte de planification.

Notamment, Martinet et al (2008) proposent une implémentation biomimétique de la carte de planification en se fondant sur l'existence des colonnes corticales (Mountcastle, 1957), interconnectées entre elles sur 6 niveaux (d'où le terme de colonne). Les colonnes, elles-mêmes subdivisées en sous-colonnes, reçoivent des entrées sensorielles et ont des connexions efférentes vers les zones sub-corticales (e.g., le striatum et le thalamus). La structure, ainsi que les connexions de ces colonnes, les rendent éligibles pour le traitement efficace des calculs complexes (Mountcastle, 1997), dont la planification (Hasselmo, 2005).

Pour Gaussier et al (2002) et Cuperlier et al (2007), parmi les cellules de lieu hippocampiques existent des cellules qui ont la particularité de représenter des transitions entre lieux (Samsonovich et McNaughton, 1997). Ces cellules de transitions sont définies dans le modèle

par deux cellules de lieux successives. Sur la base de ces cellules de transitions est construite par apprentissage déclaratif une carte de planification permettant les opérations de planification. L'association d'une cellule de la carte de planification peut ensuite être associée à une motivation (un but) par apprentissage hebbien.

Mécanisme de planification. Dans les modèles de Filliat et Meyer (2002), Martinet et al (2008) et Cuperlier et al (2007), la planification est ensuite réalisée par un mécanisme d'activation-diffusion (Burnod, 1991) : chaque nœud contient une valeur qui sera mise à 1 pour le nœud but et progressivement atténuée au fur et à mesure que les nœuds sont éloignés du nœud but, en terme de nombre de nœuds. L'animat n'aura ainsi qu'à choisir le nœud adjacent qui maximise cette valeur pour rejoindre le but.

Le modèle de Quoy et al (2002) adapte une stratégie de planification à plusieurs motivations (i.e., des besoins de différentes natures, comme la faim ou la soif seront associés à des buts distincts). Le but est de simuler une expérience dans laquelle l'animat a accès à plusieurs ressources, de type "eau" et "nourriture". Certaines ressources des deux types sont proches et il est alors plus intéressant pour l'animat d'aller vers cet emplacement. Les auteurs associent à chaque nœud du graphe de planification une activité dépendant de la motivation de l'animat (faim ou soif), cette activité est maximale lorsque ce nœud est situé à l'emplacement d'une ressource désirée. Cette activité est ensuite diffusée dans le graphe et l'animat n'a qu'à remonter le gradient pour atteindre la ressource. Un apprentissage est effectué sur les transitions entre lieux de manière à renforcer les liens les plus souvent utilisés. Les auteurs montrent que l'animat arrive à exprimer une stratégie "intelligente" à partir de ce comportement simple comme préférer les lieux où les deux types de ressources se trouvent, grâce à la dynamique instaurée par le couplage entre les motivations et l'apprentissage qui en résulte.

3.2.3 Modèles de sélection entre stratégies

Les modèles présentés ici sont fondés sur l'hypothèse de systèmes neuraux séparés et apprenant en parallèle. Ils sont donc dotés au minimum de deux modules spécialisés pour un type de stratégie et proposent une modélisation de leurs interactions.

3.2.3.1 Sélection entre modules à apprentissages associatifs hétérogènes

Uchibe et Doya (2005) proposent un modèle de sélection entre différents modules d'apprentissage spatiaux (Figure 3.3). Bien qu'hétérogènes du point de vue de leurs algorithmes d'apprentissage (certains modules dépendent d'une représentation spatiale, d'autres non), ces algorithmes sont cependant tous dépendants de l'apprentissage par renforcement. Le critère de sélection est basé sur la valeur d'état de chaque module, le module ayant la valeur d'état la plus élevée étant celui qui a la probabilité la plus forte d'être sélectionné. Les différents modules peuvent apprendre des actions générées par d'autres modules, tant que leur apprentissage est dit "*off-policy*", i.e., que son apprentissage est garanti quelle que soit l'action réalisée – à l'inverse des apprentissage "*on-policy*", où l'apprentissage nécessite l'exécution de l'action selon la politique estimée. Enfin, les auteurs ont une approche développementale, visant à optimiser l'apprentissage des différents modules, et non l'atteinte d'un but.

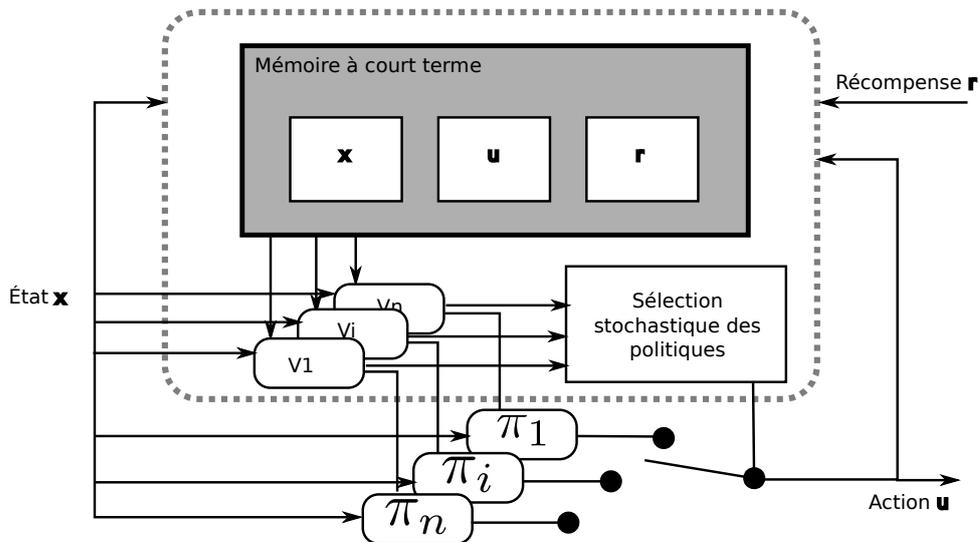


Figure 3.3 – Le modèle de Uchibe et Doya (2005). Les valeurs des différentes stratégies sont évaluées par les boîtes “ V_i ” qui sont ensuite prise en compte par le module de sélection. L’action correspondant à la politique π_i gagnante est ensuite effectuée. Les différents modules pouvant être mis à jour de manière asynchrone, le module de mémoire à court terme permet de mémoriser les triplets (x,u,r) tous les τ pas de temps, une interpolation est ensuite effectuée pour permettre une prise de décision synchrone. D’après Uchibe et Doya (2005)

Les modules sélectionnés sont au nombre de trois et, à l'exception des deux premiers qui partagent la vision du but, sont basés sur des entrées différentes :

(i) Un premier module, dit de comportement "inné", suit la direction d'un but visible tant qu'il ne rencontre pas d'obstacle et propose une direction aléatoire sinon. L'espace d'état est constitué d'un vecteur représentant une gaussienne autour de l'angle de rotation nécessaire pour rejoindre le but. Certains obstacles ne sont cependant pas observables et entraînent donc des collisions.

(ii) Le second module implémente une stratégie de guidage qui prend en compte la distance aux obstacles en plus de la vision du but, et est donc encore relativement simple. Les deux premiers modules ont donc une vision partielle de l'environnement, ce qui signifie qu'un même état peut être perçu alors même que le robot se trouve en deux endroits différents et que la politique optimale pour rejoindre le but est différente.

(iii) Enfin le dernier module est une stratégie basée sur le lieu qui prend donc en entrée un espace discret composé de 450 lieux disposés en grille, représentant la vision complète de l'environnement. L'apprentissage est ici effectué à la fois sur la politique du module et sur la connaissance de l'environnement : l'estimation de la position du robot est améliorée à chaque pas de temps, ce qui lui permet d'obtenir un modèle du monde fiable après apprentissage.

Les expériences robotiques, où un robot doit rejoindre un but fixe en évitant des obstacles de différentes tailles, montrent une nette supériorité du modèle de sélection sur les performances des stratégies seules. En sélectionnant chaque stratégie en fonction de son évaluation d'état, la sélection favorise celle qui "connait" le mieux son environnement à un instant donné. Naturellement la stratégie de lieu est sélectionnée en dernier, qui est bien plus performante que les stratégies innées et de guidage, du fait de sa connaissance complète de l'environnement.

En dépit de ses nombreuses qualités, le modèle ne s'adresse cependant pas aux problèmes suivants : que se passe-t-il en cas de conflit perceptif (e.g., vision du but différente de sa dernière position connue) ? Quelles sont ses performances en l'absence du module inné (dont la connaissance du but est immédiate) ? Comment ce modèle peut-il gérer des stratégies qui ne dépendraient pas de l'apprentissage par renforcement ?

3.2.3.2 Sélections "ad hoc" entre stratégies procédurales et déclaratives

Modèle de Guazzelli et al (1998). Ce modèle réalise une sélection entre une stratégie de type Taxon et une autre de type Planification. Les deux types de stratégies prennent en entrées les mêmes entrées sensorielles (vision des murs et amers). La stratégie Taxon permet principalement ici de représenter les mouvements possibles (théorie des *affordances*, Gibson, 1966) : ainsi, les actions possibles associées par défaut avec 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). La motivation pour un but particulier est ainsi influencée à la fois par l'état interne (e.g., soif, faim, etc.) mais également par l'environnement (e.g., si le robot a faim et qu'il perçoit la nourriture, alors la faim n'en sera que d'autant plus grande). La motivation ne baisse que lorsque l'animat atteint effectivement le but.

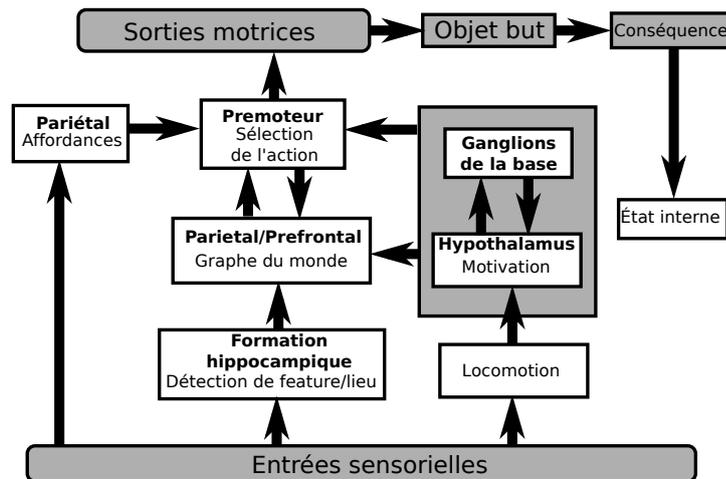


Figure 3.4 – Le modèle de Guazzelli et al (1998). Assemblage des composants : les *affordances* du pariétal, l'information motivationnelle et les stimuli associés sont combinés dans le cortex prémoteur où est sélectionné le prochain mouvement. D'après Guazzelli et al (1998)

La stratégie de lieu "WG" est une implémentation de la *World Graph Theory* proposée par Arbib et Lieblch (1977). Sa modélisation est fondée sur la construction d'un graphe représentant les relations topologiques entre les différents lieux visités, simulant les fonctions des cellules de lieu et de direction 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. Il n'y a pas d'exploration explicite, mais celle-ci est présente dans l'apprentissage par renforcement associé à la sélection de la direction finale.

Ce modèle peut effectuer une sélection entre plusieurs buts disponibles (e.g., présence d'eau et de nourriture) qui sera modulée par la motivation de l'animat, par la prise en compte d'états motivationnels fluctuants, comme la faim et la soif. La direction finale consiste en une sommation des suggestions de déplacements issues des deux stratégies de navigation passant par un *winner-takes-all*. Au final, ce modèle a été testé avec succès par la reproduction d'une expérience dans un labyrinthe radial, mais l'absence d'apprentissage au niveau de la sélection le rend peu adaptatif, la sélection finale de l'orientation à prendre ne dépendant que de la motivation de l'animat et des directions possibles.

Modèle de Girard et al (2005). Ce modèle s'intéresse également aux aspects motivationnels de la sélection de stratégie dans une tâche dans laquelle l'animat doit survivre le plus longtemps possible en "consommant" de l'énergie "potentielle" dans certaines zones de l'environnement et en la "digérant" dans d'autres pour la transformer en énergie "effective". Deux stratégies sont implémentées et partagent les mêmes entrées sensorielles : une stratégie de Planification (Plan)(correspondant au modèle de Filliat et Meyer, 2003) et une stratégie Taxon (Tax, correspond en l'orientation fournies par Prox(res) des différents objets présents dans l'environnement,

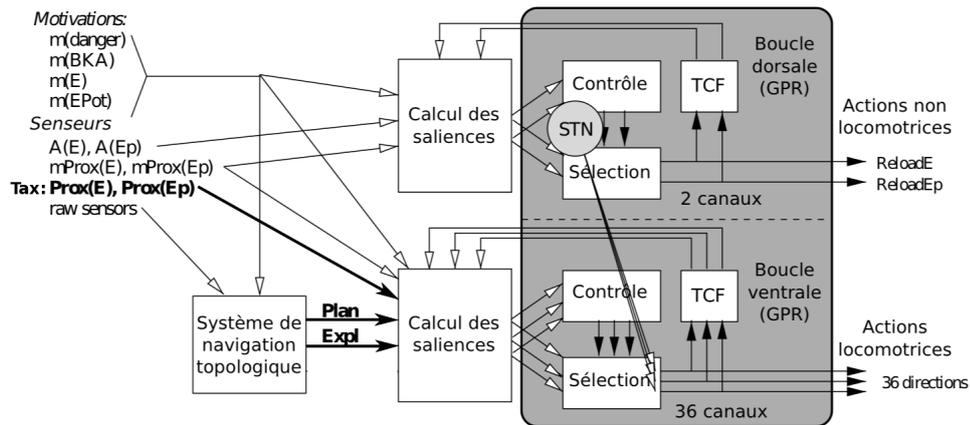


Figure 3.5 – Le modèle de Girard et al (2005). Les saillances de stratégies (Tax, Plan) sont évaluées selon les senseurs et les motivations, puis l'action finale est sélectionnée. TCF : *Thalamo-cortical feedback*, ou retour des boucles thalamo-corticales, STN : Substance noire, GPR : le modèle des ganglions de la base de Gurney et al (2001a,b).

Figure 3.5). Chaque stratégie propose une direction pour chaque but (zones "Consommation" et "Digestion") dont la valeur est modulée par la saillance du but recherché, en fonction d'états motivationnels fluctuants, comme le besoin de consommer, de digérer ou de rafraîchir sa carte cognitive. Les motivations sont représentées sous forme d'états internes et sont ordonnées : par exemple, la faim conditionnera la recherche de nourriture puis, lorsqu'elle est atteinte, la motivation sera de digérer cette nourriture dans un endroit sûr. Une stratégie d'exploration (Expl) est également présente, sélectionnée de la même manière que les deux stratégies.

Le choix de la direction finale est géré par un module spécifique (simulant la boucle cortico-basale ventrale) qui calcule la somme pondérée des deux propositions de direction. Le système modélise très en détail cette boucle cortico-basale (Figure 3.5), ce qui le rend plus biomimétique que le modèle de Guazzelli et al (1998) avec lequel il partage le principe de sélection. Il montre des capacités de survie raisonnables (avec les deux stratégies sélectionnées par le système, l'animat survit trois fois plus longtemps qu'avec le Taxon seul) dans un environnement complexe où l'animat doit utiliser les différentes stratégies. Cependant, le critère de sélection n'est pas appris, et les poids pour chacune des stratégies doivent donc être préparés à l'avance par l'expérimentateur selon l'environnement. Par exemple, lorsque l'animat a planifié un chemin vers une ressource invisible, il ne pourra rejoindre une ressource visible se trouvant sur son chemin que par un jeu de paramètres choisi a priori pour pondérer beaucoup plus une stratégie Taxon qu'une stratégie Planification.

3.2.3.3 Sélections adaptatives entre stratégies procédurales et/ou déclaratives

Sélection bayésienne dans un environnement discret (Daw et al, 2005). Bien que non adapté spécifiquement à la navigation, mais au conditionnement opérant, le modèle de Daw

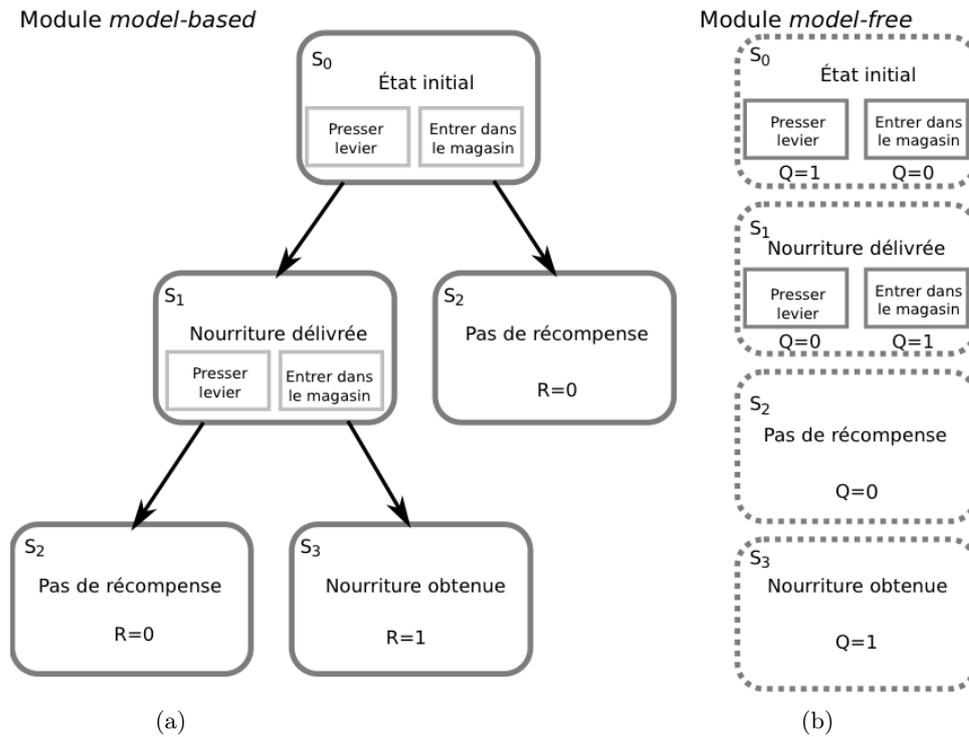


Figure 3.6 – Le modèle de Daw et al (2005) simule une expérience classique de conditionnement opérant (voir texte). (a) Le module *model-based* permet un accès à l'ensemble du monde, autorisant ainsi la mise à jour quasi-instantanée des valeurs d'action dans l'état S_0 dès qu'un état final est atteint. (b) En revanche, le module *model-free* ne possède pas de telle représentation, mais plusieurs états non connectés entre eux. L'apprentissage par Q-learning se fait alors progressivement et est donc plus lent que celui du module *model-based*.

et al (2005) a été développé pour permettre la sélection entre deux systèmes d'apprentissages distincts, indépendant et dépendant de l'hippocampe : le premier consiste en une association simple entre l'état courant et les actions du modèle (module *model-free*, appris par *Q-learning*) tandis que le second possède un "modèle du monde", une représentation de l'environnement sous forme d'un arbre dans lequel les transitions états-actions sont représentés (module *model-based*). Le premier est un système d'apprentissage S-R tandis que le second pourrait implémenter un graphe représentant une carte topologique. L'exploration est gérée par un *softmax* (cf. section 3.2.1.1).

Pour sélectionner entre ces deux systèmes, les auteurs proposent d'estimer l'erreur cumulative de différence temporelle de ces deux systèmes – en utilisant une approximation bayésienne de ces systèmes et en calculant ensuite l'estimation a priori de leur incertitude – et d'opérer un changement de système lorsque l'erreur du système actuellement utilisé dépasse un certain seuil. D'après les auteurs, cette sélection pourrait s'effectuer dans le mPFC ou le DMS.

Les auteurs ont adapté ce modèle à une expérience simple de prise de décision : un rat est déposé dans une cage dans laquelle se trouve un magasin de nourriture et un levier situé à côté du magasin. Le nombre d'états et d'actions à représenter est donc particulièrement faible (4 états, cf. Figure 3.6). Pour obtenir de la nourriture, le rat doit presser le levier et se diriger ensuite vers le magasin afin de la consommer. Dans cette expérience, si l'on empêche la livraison de nourriture tôt dans l'apprentissage, le rat arrête rapidement d'appuyer sur le levier. En revanche, après un long entraînement, le rat persiste à appuyer sur le levier longtemps après l'arrêt de livraison de nourriture, d'où l'hypothèse d'un changement au cours de l'expérience entre un système d'apprentissage rapide et flexible (*goal-directed*) et un autre, plus lent et inflexible (*habit*)

Ici, le modèle permet de reproduire le passage *goal-directed* vers *habits* observé, et donc ici le passage du système *model-based* vers le système *model-free*. Les auteurs font ici l'hypothèse que le système *model-based* est plus incertain (i.e., l'erreur cumulée augmente plus rapidement) que le système *model-free*. L'expérience reproduite étant trop simple pour que le modèle du monde produise de lui-même cette incertitude (i.e., le passage d'un état à l'autre n'est pas garanti à 100%), ils y ajoutent un bruit ad-hoc.

Ainsi, bien que le critère de sélection basé sur l'incertitude soit plausible, se pose le problème de l'origine de ce calcul d'incertitude qui n'est pas réglée par le présent modèle. De fait se pose également la question du passage à l'échelle : l'algorithme utilisé pour l'apprentissage *model-based* est extrêmement coûteux, puisqu'il nécessite la construction et la liaison de l'ensemble des états possibles pour l'expérience. L'expérience reproduite ne contient que 4 états, mais si l'on veut adapter ce modèle à la navigation (où de telles transitions ont été observées, cf. Chapitre 2), et plus précisément à un espace d'états continu, on se retrouve face à une explosion combinatoire du nombre d'états, ce qui rend l'algorithme *model-based* totalement inutilisable.

Sélection par champs d'attracteurs dans un environnement continu (Lagarde, 2010).

Dans ce récent travail, deux stratégies sont proposées pour apprendre une même trajectoire dans le contexte d'une navigation visuelle. La première est spatiale et se fonde sur la notion de cellules de lieu et des perceptions et mouvements associés à des zones distinctes de l'environnement. Elle informe sur l'endroit où doit se réaliser l'action. La deuxième est temporelle et est liée à l'apprentissage d'une séquence temporelle de mouvements pendant la navigation. Elle informe sur quand réaliser l'action. Ces deux stratégies sont ensuite fusionnées et exécutées en parallèle. La fusion des réponses délivrées par les deux stratégies est réalisée avec l'utilisation de champs de neurones dynamiques, qui infèrent une sélection des actions par la convergence vers attracteur le plus proche. Ces deux stratégies peuvent ainsi opérer en parallèle ou coopérer. Ce principe peut difficilement encore s'appliquer à la confrontation de stratégies de guidage et de lieu.

Sélection associative dans un environnement continu (Chavarrriaga et al, 2005).

A l'instar du modèle de Daw et al (2005), c'est ici une valeur interne commune aux stratégies de navigation qui est comparée pour la sélection. Son critère de sélection peut d'ailleurs être vu comme une adaptation à l'espace continu de celui de Daw et al (2005). Deux experts, Locale et Taxon, reçoivent des entrées séparées (resp. amers distaux et proximaux) et apprennent tous les deux à leur associer des actions d'orientation par une variante de Q-learning (utilisant l'algorithme de Strösslin et al, 2005, cf. section 3.2.2.2), ce qui leur permet de mettre simultanément

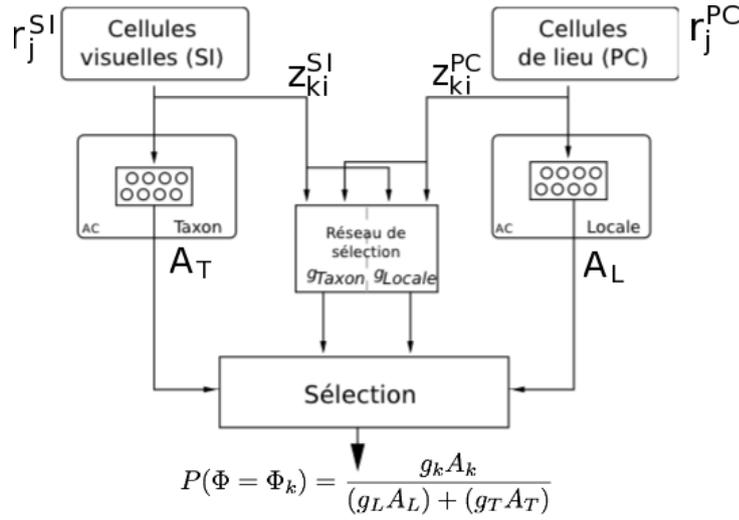


Figure 3.7 – Le modèle de Chavarriaga et al (2005).

à jour leur erreur respective de différence temporelle (Figure 3.7). L'exploration émerge de la sélection quasi-aléatoire au début des deux stratégies.

C'est cette erreur qui va pondérer la mise à jour des poids z_{ki}^{PC} et z_{ki}^{SI} du réseau de sélection liant les entrées des différents experts ("cellules visuelles" pour le Taxon et "cellules de lieu" pour le Local, construites à partir de l'estimation des distances de l'animat aux murs) à des valeurs de sélection (*gating-values* g_k), dont la valeur représente la confiance des experts pour trouver le but.

En effet, un module de sélection indépendant calcule les valeurs suivantes :

$$g_k = \sum_{j \in PC} (z_{kj}^{PC} r_j^{PC}) + \sum_{j \in SI} (z_{kj}^{SI} r_j^{SI}), \quad (3.13)$$

où r_j^{PC} et r_j^{SI} correspondent respectivement à la valeur d'activation de la j^{eme} cellule de lieu (*Place Cells*) et de la j^{eme} cellule visuelle (*Sensory Input*).

Ainsi plus un expert aura une erreur TD basse, plus sa valeur de sélection sera élevée. La sélection finale de l'expert pondère cette valeur de sélection avec la prédiction de récompense des experts, autorisant ainsi un expert ayant une valeur de sélection basse à être sélectionné, s'il promet une grande récompense :

$$P(\Phi = \Phi_k) = \frac{g_k A_k}{(g_L A_L) + (g_T A_T)}, \quad (3.14)$$

où $P(\Phi = \Phi_k)$ est la probabilité que la prochaine direction de mouvement soit celle proposée par l'expert k , ($k \in (T, L)$), g_k la valeur de sélection de l'expert k et A_k sa prédiction de récompense.

On retrouve dans ce modèle l'idée de sélection sur la base des performances déjà défendue par l'hypothèse du comparateur (section 3.1.2). La règle de mise à jour des poids n'est toutefois

pas son équivalent, mais peut être modifié dans cette optique.

Ce modèle a pu reproduire deux expériences dans lesquelles les deux stratégies interagissaient de façon compétitive ou coopérative (e.g., expérience de Pearce et al (1998), par Chavarriaga et al, 2005; Dollé et al, 2008). Ces expériences n’abordaient toutefois pas les questions de la flexibilité de la stratégie de “place” (Locale ou Planification) ni les phénomènes associatifs (blocage, occultation, etc.).

3.2.4 Avantages et limites des modèles de sélection de stratégies

Les modèles computationnels dévolus à la fois à la navigation et aux interactions entre systèmes de mémoires sont complémentaires, mais ils ne répondent chacun qu’à une partie du rôle de la sélection de stratégies de navigation : les modèles permettant une sélection entre systèmes appris différemment ne sont pas adaptatifs (Guazzelli et al, 1998; Girard et al, 2005) ou sont conçus dans des environnements discrets ne supportant pas le passage à l’échelle nécessaire en navigation (Daw et al, 2005). Enfin, d’autres modèles proposant un critère adaptatif ne peuvent être généralisés à des systèmes appris différemment (Chavarriaga et al, 2005; Lagarde, 2010). Certains modèles proposent toutefois une implémentation biomimétique qui leur permet d’effectuer de prédictions, par exemple les conséquences de la lésion d’un système.

Nous retiendrons toutefois dans Guazzelli et al (1998) ou Girard et al (2005) la prise en compte de plusieurs systèmes motivationnels, indispensable pour une navigation vraiment autonome.

3.3 Conclusion du chapitre : vers un nouveau modèle de sélection de stratégies de navigation

Nous avons vu ici que les modèles associatifs de compétition entre amers peuvent reproduire de nombreux phénomènes – comme le blocage ou l’occultation d’amers –, mais ne donnent guère d’informations sur les systèmes neuraux engagés. Ils suggèrent toutefois des pistes sur la nature possible de leurs interactions (e.g., compétition au niveau de l’apprentissage ou au niveau de leur sélection). Le principe de compétition entre amers doit donc être pris en compte pour un futur modèle. Par ailleurs un modèle de sélection adaptative entre systèmes parallèles a pu rendre compte de compétition et coopération entre stratégies. Il peut également apporter des informations sur les structures nerveuses concernées, en mettant en œuvre les systèmes isolément ou en interaction.

En vue de reproduire les différents effets de l’apprentissage spatial décrits au Chapitre précédent, l’objectif est de concevoir un modèle qui prenne à la fois en compte :

- (i) l’hypothèse de compétition entre amers
- (ii) le principe de systèmes de navigation apprenant en parallèle, par des mécanismes qui peuvent être différents
- (iii) un module de sélection adaptatif, ayant un mécanisme d’apprentissage indépendant des systèmes précédents

Le modèle proposé est inspiré de celui de Chavarriaga et al (2005). Bien que ne pouvant pas

sélectionner entre des stratégies différemment apprises, son critère de sélection est relativement proche de celui proposé par Daw et al (2005). Ce critère est sous-tendu par un apprentissage par renforcement, qui constitue l'implémentation biomimétique la plus proche de la théorie associative. Il peut donc constituer une base pour reproduire les phénomènes de compétition entre amers.

Ce modèle va en outre être conçu pour permettre à chaque système de navigation d'apprendre soit sur la base de certains amers (e.g., amers proximaux pour stratégies de guidage, amers distaux pour stratégies de lieu), soit sur la base de tous les amers. En effet, bien que certaines expérimentations animales indiquent que chacun prendraient en compte des amers spécifiques, nous estimons que certaines conditions environnementales pourraient contraindre les animaux à tenir compte des deux types d'indices environnementaux.

Troisième partie

Méthode

Chapitre 4

Description du modèle

Les stratégies mises en œuvre dans le modèle sont décrites selon la nature de leurs entrées et l'apprentissage de leurs sorties. Nous décrirons ensuite le critère et le mécanisme de sélection de stratégie, qui constituent le principal apport de ce travail. Le Tableau complet des paramètres se trouve dans l'annexe B (Tableau B.1).

Pour la compréhension de ce chapitre, nous décrivons brièvement les caractéristiques de l'animat, le dispositif utilisé et la tâche générale. L'animat utilisé est circulaire, de 10cm de diamètre. Il possède une vue panoramique, (360 degrés, distance minimale de 5 cm, maximale dépendant de la taille de l'amer, voir 4.1.1.1), une vitesse constante de 6 cm par pas de temps, et de possibilité d'orientations illimitées. Il s'agit donc d'un robot simulé idéal dont les simplifications volontaires sensorielles et motrices nous permettront de nous focaliser sur l'objectif du modèle et non des problèmes attendant à son implémentation robotique. Le dispositif commun est la piscine de Morris. Le diamètre de la plateforme, sa visibilité, le nombre et les amers utilisés varieront selon les expériences. La tâche sera toujours de rechercher à atteindre la plateforme, ce qui correspond à un seul but, dépendant d'un seul système motivationnel.

Tableau 4.1 – Experts de navigation selon leurs entrées, leurs apprentissages et leurs référentiels. Proc=apprentissage par renforcement procédural, Decl=apprentissage par recherche de graphe déclaratif.

Expert	Entrées	Apprentissage	Référentiel
Exploration	\emptyset	\emptyset (aléatoire)	Egocentrique
Taxon	Vision	Proc	Egocentrique
Direction	Vision	Proc	Allocentrique
Local	Cellules de lieux	Proc	Allocentrique
Planification	Carte topologique	Decl	Allocentrique

4.1 Codage des experts

Les stratégies de navigation sont gérées par des experts séparés dont les entrées, l'algorithme d'apprentissage et le référentiel peuvent être différents. Le Tableau 4.1 résume l'ensemble des experts qui seront utilisés par la suite selon ces trois critères.

Tous les experts offrent en sortie une action d'orientation entre 0 et 2π . Nous décrivons d'abord les stratégies de guidage (Taxon, Direction) puis les stratégies associées à l'information de lieu (Locale, Planification), et enfin la stratégie d'exploration (étant absente dans les autres stratégies).

4.1.1 Deux experts de guidage procéduraux : les experts Taxon et Direction

4.1.1.1 Entrées visuelles

Un stimulus visuel est encodé par un vecteur de N_{SC} cellules "sensorielles" (*Sensory Cells* SC), représentant une vue panoramique de l'environnement en niveaux de gris. En fait, ces cellules pourraient plutôt être nommées "cellules d'amer", car l'activité d'une cellule i signale le niveau de perception d'un seul amer dans la direction Φ^i de l'animal simulé :

$$r_i^{SC} = \exp\left(-\frac{(\Phi^l - \Phi^i)^2}{2\sigma_{SC}^2}\right), \quad (4.1)$$

Où Φ^l est la direction de l'animat à l'amer. La largeur de la gaussienne σ_{SC} centrée sur la direction de la cible augmente lorsque l'animat s'en rapproche, simulant ainsi l'augmentation de la taille de l'amer dans son champ visuel. Pour ce faire, l'écart-type σ_{SC} est obtenu à partir de la largeur à mi-hauteur de la gaussienne, égale au diamètre D_r de la projection de l'amer sur la rétine (Figure 4.1a) :

$$\sigma_{SC} = \frac{D_r}{2\sqrt{2\log 2}} \quad (4.2)$$

avec :

$$D_r = \frac{D_a}{\Delta_{R \rightarrow L}} R_r, \quad (4.3)$$

où D_a est le diamètre réel de l'amer et R_r le rayon de la rétine (fixé à 0.5 cm) et $\Delta_{R \rightarrow L}$ la distance de l'animat à l'amer.

De fait la perception d'un amer par l'animat sera conditionnée par la taille de cet amer, sa distance, et le nombre de SC codant pour la représentation : Pour 100 SC, un amer de 50 cm de diamètre sera perçu jusqu'à 7,50 m. Son diamètre sur la rétine sera alors de 0.3 cm, ce qui correspond à la zone d'activation d'une SC. Nous ne prétendons pas ici simuler précisément la vision des animaux : nos simulations se restreignent à la piscine de Morris, dans lesquels les amers proximaux et distaux sont considérés comme perçus par les rats. Notre modèle est donc suffisant pour simuler les distances et différentes tailles des amers, et ainsi l'importance relative de chacune d'entre elles. En l'absence d'amer, les entrées sont constituées d'un bruit uniforme de moyenne égale à 0.1.

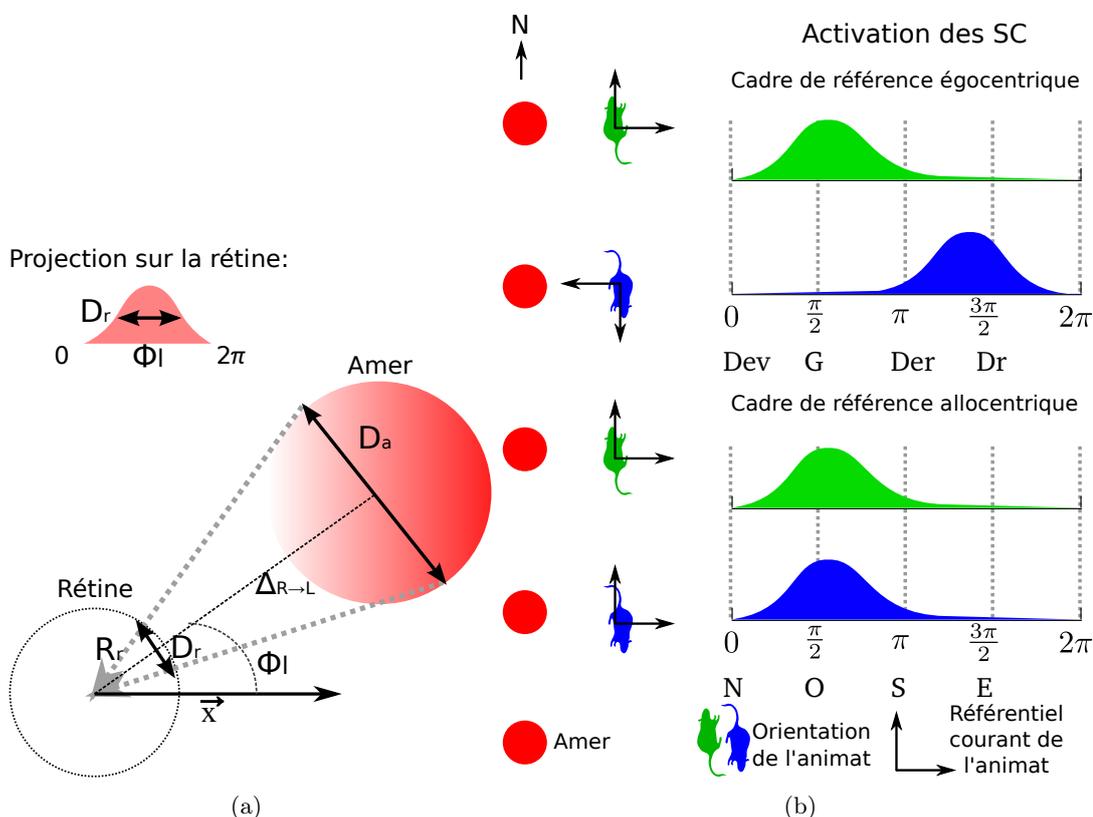


Figure 4.1 – (a) Projection de la vision d'un amer sur la rétine. Voir texte pour les équations liant les variables. (b) Représentation interne de l'amer (disque rouge) situé à l'Ouest de l'animat dans les deux référentiels : dans un référentiel égocentrique (haut), l'amer perçu par l'animat, orienté vers le nord et le sud, sera représenté par des cellules sensorielles fortement activées dans les directions $\pi/2$ et $3\pi/2$. Dans un référentiel allocentrique (bas), l'amer sera représenté par des cellules actives dans la direction $\pi/2$, quelle que soit l'orientation de l'animat (Dev : Devant ; G : Gauche ; Der : Derrière ; Dr : Droite ; N : Nord ; O : Ouest ; S : Sud ; E : Est).

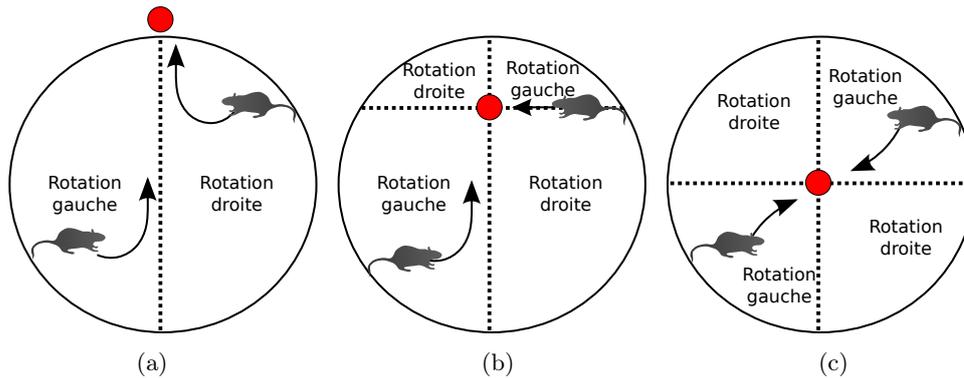


Figure 4.2 – Lorsqu’un seul amer est présent (disque rouge), l’animat peut quand même créer un référentiel allocentrique en fonction de ses distances à l’amer et aux parois de la piscine, et de la rotation qu’il doit effectuer pour se positionner dans sa direction. Par exemple, dans les cas indiqués par les deux premières figures, l’amer est à la même place sur le champ visuel pour les deux positions et orientations indiquées, mais la rotation pour se diriger vers l’amer n’est pas la même que l’amer soit distal (a) ou proximal (b). Le seul cas problématique est quand l’amer se trouve au centre de la piscine (c), car la distance aux parois est non discriminante).

Le référentiel dans lequel l’amer sera ensuite perçu pourra être égo- ou allocentré (cf. paragraphe ci-dessous et Figure 4.1 b). Pour l’expert Taxon, il sera toujours égocentré, pour l’expert Direction, toujours allocentré.

4.1.1.2 Apprentissage par renforcement procédural

Les réponses motrices des experts de guidage devant être apprises sont encodées par $N_{AC} = 36$ cellules d’actions (*Action Cells AC*). Chaque AC reçoit les entrées de toutes les SCs et code une direction de mouvement $\phi_i = \frac{2\pi i}{N_{AC}}$.

Ces experts peuvent donc fonctionner dans deux référentiels différents. Ils influent principalement sur l’angle utilisé pour calculer l’emplacement de l’amer dans le champ visuel de l’animat en entrée de l’expert et les actions d’orientation en sortie.

- **Référentiel égocentrique** : La direction Φ^l de l’amer perçu est calculée comme l’angle entre l’orientation courante de l’animat et la position de l’amer par rapport à cette orientation.
- **Référentiel allocentrique** : Il est intégré en calculant la direction Φ^l comme l’angle formé par la position de l’amer et un axe absolu \vec{x} , indépendant de l’orientation de l’animat. A l’exception du cas où le but est visible, nous faisons l’hypothèse que la présence d’indices, qu’ils soient distaux ou proximaux, permet en effet la construction d’un référentiel allocentrique. Lorsqu’un seul amer est présent, la construction d’un tel référentiel est possible par l’inférence de la position de cet amer avec la distance aux parois de l’environnement (Figure 4.2). Il s’agit évidemment ici d’une simplification de modélisation, qui

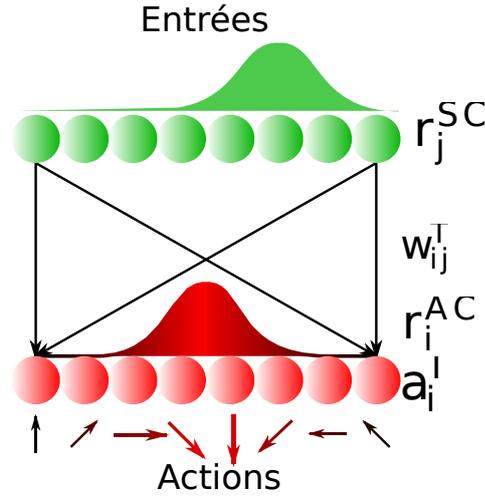


Figure 4.3 – Le réseau de neurones modélisant la structure des experts de guidage. Voir texte pour le détail des variables.

ne traduit pas la difficulté de construire un tel référentiel qui variera selon la nature et le nombre d'amers présents.

Nous présentons d'abord l'apprentissage d'un expert par rapport à un seul amer. Les modifications pour traiter plusieurs amers seront détaillées dans la section suivante.

L'apprentissage "Stimulus-Réponse" des experts de guidage consiste en une adaptation de l'algorithme de Q-learning aux espaces d'états et d'actions continus. L'expert prend donc la forme d'un réseau de neurones *feed-forward* liant les entrées SC aux actions d'orientations (Figure 4.3).

L'activité des ACs représente la valeur (i.e., récompense espérée) d'un mouvement dans la direction correspondante et est calculée comme suit :

$$a_i^T(t) = \sum_{j=1}^{N_{SC}} r_j^{SC}(t) w_{ij}^T(t). \quad (4.4)$$

L'activité $a_i^T(t)$ d'une AC i est donc la Q-valeur du couple état-action correspondant. L'activité totale de la population est interprétée comme un code de population pour la direction Φ^T pour le prochain mouvement de l'animat, proposée par l'expert Taxon ou Direction (Strösslin et al, 2005) :

$$\Phi^T(t) = \arctan\left(\frac{\sum_i a_i^T(t) \sin(\phi_i^T)}{\sum_i a_i^T(t) \cos(\phi_i^T)}\right). \quad (4.5)$$

La Q-valeur $A^T(\Phi^T(t))$ associée à $\Phi^T(t)$ est définie comme l'interpolation des deux a_i^T codant pour les directions les plus proches de $\Phi^T(t)$. Les associations SR sont encodées dans les poids w_{ij}^T , appris au moyen d'un algorithme de Q-learning (Sutton et Barto, 1998) :

$$\Delta w_{ij}^T = \eta^T \delta^T(t) e_{ij}^T(t+1). \quad (4.6)$$

Ici, η^T est le taux d'apprentissage et $\delta^T(t)$ est l'erreur de prédiction de récompense :

$$\delta^T(t) = R(t+1) + \gamma^T A^T(\Phi^T(t+1)) - A^T(\Phi^T(t)), \quad (4.7)$$

où $R(t+1)$ est la récompense délivrée au temps $t+1$, $0 < \gamma < 1$ le facteur d'atténuation de la prochaine récompense (Sutton et Barto, 1998), et A^T la Q-valeur de l'action exécutée au temps t , estimée par l'expert.

Le calcul de l'erreur de prédiction de récompense et la mise à jour des poids correspondants sont aussi effectués lorsque l'action au temps t n'a pas été générée en suivant la politique de l'expert, tout en lui permettant d'apprendre. Les traces d'éligibilité e_{ij}^T accélèrent l'apprentissage en mémorisant les couples état-action expérimentés dans le passé :

$$e_{ij}^T(t+1) = r_j^{SC}(t)r_i^{AC}(t) + \lambda^T e_{ij}^T(t), \quad (4.8)$$

où λ^T est le taux d'atténuation de la trace et r_i^{AC} correspond à :

$$r_i^{AC}(t) = \exp\left(\frac{-(\phi_i^T - \Phi^T(t))}{(2\sigma^T)^2}\right), \quad (4.9)$$

Ce terme représente l'activité de la cellule d'action dans la phase de généralisation (Strösslin et al, 2005) et permet aux actions proches de celle effectuée de mettre à jour leurs poids dans la même direction. L'utilisation d'une telle phase de généralisation, ainsi que l'équation 4.5 permettent à la stratégie procédurale d'opérer dans un espace d'états et d'actions continu. La conversion des activités de la population AC depuis leur forme initiale à la forme généralisée consiste en une gaussienne autour du centre de masse et peut donc être implémentée de manière biologiquement plausible par un réseau dynamique récurrent (Wilson et Cowan, 1972; Denève et al, 1999), ici non modélisé pour des raisons de simplicité.

Selon la nature des entrées, le référentiel des sorties motrices peut être soit égocentrique, soit allocentrique. Dans le premier cas (expert Taxon), la direction zéro correspond à l'orientation courante de l'animat. Dans le second cas (expert Direction), toutes les directions sont données en fonction d'une direction zéro fixée arbitrairement dans ce référentiel, indépendamment de l'orientation de l'animat. Dans le cas où une contradiction apparaît entre les différents amers (par exemple, par la rotation des amers proximaux ou distaux), notre référentiel sera ancré, de manière ad hoc, sur l'amer le plus éloigné, celui-ci étant a priori le plus stable. Cet expert permet d'atteindre le but en suivant une direction générale en fonction de la présence d'amers, distaux ou proximaux, mais ne code pas pour un lieu précis et, par la nature de son apprentissage procédural, s'avère peu flexible.

4.1.2 Un expert de lieu procédural : l'expert Local

Nous proposons ici une implémentation de la stratégie d'*action associée à la reconnaissance de lieu* (Trullier et al, 1997) : elle consiste à associer à un lieu (ici représenté par une activation neurale unique) une action d'orientation afin de rejoindre un but donné, sans que l'information de transition entre deux lieux ne soit codée, ni par les cellules de lieu, ni par la stratégie. Il est donc impossible, en un lieu donné, d'inférer une trajectoire permettant de rejoindre tout autre lieu.

4.1.2.1 Entrées

Les entrées sont représentées par une carte de cellules de lieux. Nous avons intégré un modèle d'hippocampe permettant l'apprentissage de ces cellules dont l'activation est, entre autres, dépendante de l'environnement.

Le modèle que nous utilisons a été développé par Ujfalussy et al (2008). Il apporte aux modèles existants (Arleo et Gerstner, 2000) la prise en compte des cellules de grilles, récemment découvertes chez le rat (Hafting et al, 2005). Ces dernières, situées dans le cortex entorhinal médial, présentent la particularité d'avoir une activation topographique : Chaque cellule est réceptive à une échelle spatiale et une orientation particulières, ce qui leur donne une organisation topographique triangulaire (Figure 4.4 b, haut).

A l'origine, le modèle est composé de trois parties, simulant chacun une structure neurale précise (cortex entorhinal pour les cellules de grilles, gyrus denté pour les cellules de lieu et CA3 pour la mémoire auto-associative). Néanmoins, pour nos besoins, seuls le cortex entorhinal et le gyrus denté, dont nous présentons ici la modélisation, nous sont nécessaires, le gyrus denté fournissant des cellules de lieu exploitables (Figure 4.4 a). De fait, dans ce modèle, le CA3 a surtout un rôle de consolidation de la mémoire qui permet d'obtenir une information de lieu par la reconnaissance partielle des informations environnementale, ce dont nous pouvons nous passer ici. Pour une description du modèle complet, on pourra se référer à l'article d'Ujfalussy et al (2008).

Ce modèle est de nature "apprentissage et rappel", ce qui signifie que le modèle est exclusivement activé soit en mode apprentissage, où la représentation spatiale est apprise, soit en mode rappel, où la représentation spatiale est rappelée et peut fournir une activation des cellules de lieu cohérente avec la position de l'animat dans son environnement. Ces deux modes ne peuvent être activés simultanément, le mode apprentissage devant naturellement précéder le mode rappel.

Cortex entorhinal (Cellules de grilles). La modélisation du cortex entorhinal (EC) est basée sur les cellules de grille. Le taux d'activation d'une cellule peut prendre une valeur de l'ensemble $G(i) = \{0.0, 0.0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 1.0, 1.0\}$ où i est donné par :

$$i(\mathbf{r}) = 20 \left(\sqrt{\frac{1}{3} \sum_k \cos^2(\omega_k \cdot (\frac{\mathbf{r}}{\zeta} - c))} + 0.5 \right), \quad (4.10)$$

\mathbf{r} est la position de l'animat, ζ fixe l'échelle, ω_k l'orientation et c la phase de la grille. Ce calcul peut être implémenté avec les informations proprioceptives, plutôt que les informations de position, de manière plus réaliste et plausible en ajoutant un feed-back entre les cellules des lieux et les cellules de grilles afin de stabiliser leur activation (Samu et al, 2009).

Comme dans le modèle original, 5 échelles différentes ($\{36, 41, 47, 53, 61\}$ centimètres), 5 phases différentes (entre 0 et π) et 3 orientations différentes (entre 0 et $2\pi/3$) sont utilisées, ce qui fait au final 75 cellules de grilles, chacune spécialisée pour une orientation, une échelle et une phase différentes. Aux cellules de grilles sont ajoutées la vision panoramique de l'environnement, sous forme d'un vecteur de 100 niveaux de gris codant pour chaque orientation. La vision d'un amer est calculée comme pour les entrées des experts de guidage, mais les représentations des différents amers sont sommés plutôt que concaténées, favorisant ainsi la formation d'un *snap-*

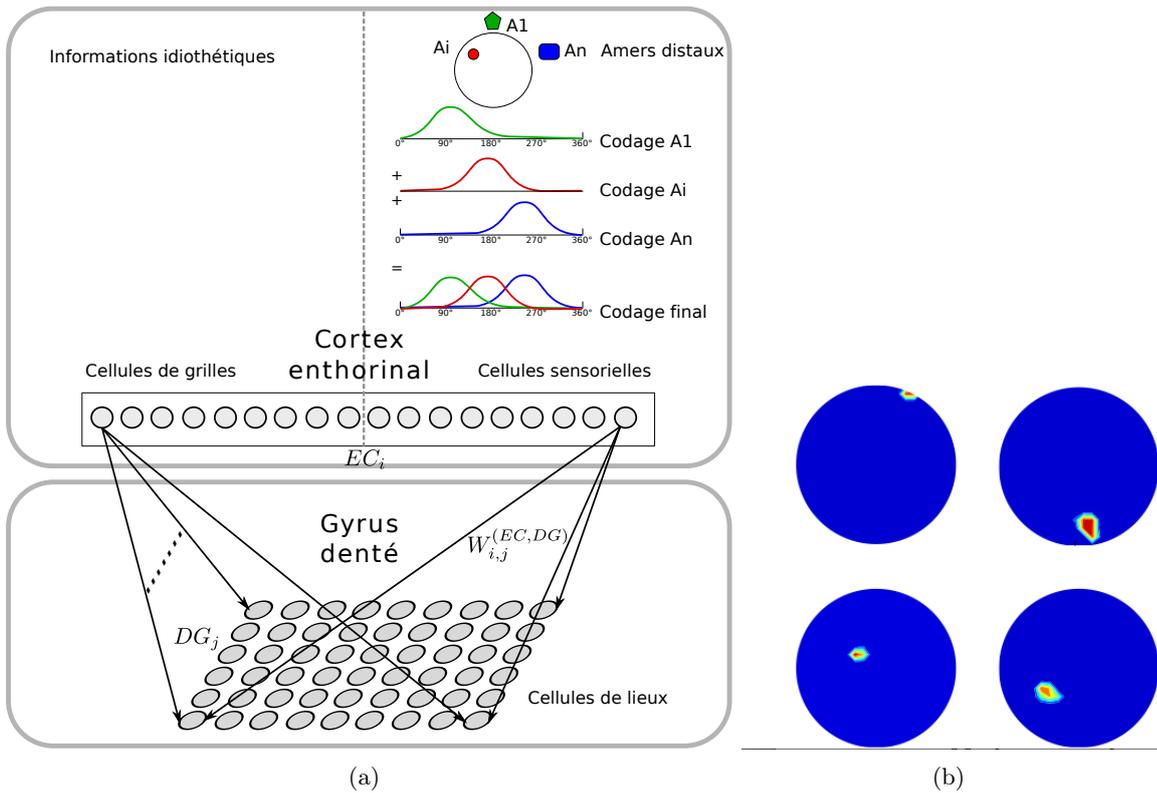


Figure 4.4 – (a) Le modèle d’hippocampe utilisé : Le cortex entorhinal encode les informations idiothétiques (cellules de grilles, à gauche) et la vision de l’environnement, représentée par la somme des codages des différents amers. Le gyrus denté reçoit les sorties du EC et réalise un apprentissage hebbien sur les poids pour construire les cellules de lieu. (b) Champs d’activation de quatre cellules de lieu apprises, choisies aléatoirement.

shot de l'environnement (Figure 4.4 a).

Gyrus denté (Cellules de lieu). Dans ce modèle, les cellules de lieu du gyrus denté (DG) traduisent par un encodage épars l'activation des cellules du cortex entorhinal. L'activation des cellules du DG est calculée par la somme des entrées du EC pondérée par les poids synaptiques :

$$h_j^{EC \rightarrow DG} = \sum_i W_{i,j}^{(EC,DG)} EC_i, \quad (4.11)$$

Une fonction non-linéaire d'activation f_j est ensuite appliquée pour calculer le taux d'activation de chaque cellule :

$$DG_j = f_j(h^{EC \rightarrow DG}, s^{DG}), \quad (4.12)$$

où s^{DG} indique le niveau de "rareté" (*sparseness*) de l'activation en sortie. Cette fonction $f(x)$ permet de ne retenir que les activations des cellules les plus activées et assigne une valeur nulle aux autres (Figure 4.4 b). Les valeurs d'activités du vecteur x les plus basses sont mises incrémentalement à zéros jusqu'à que la condition suivante soit remplie :

$$\frac{\langle x \rangle^2}{\langle x^2 \rangle} \leq s^{DG} \quad (4.13)$$

A l'instar du EC, l'apprentissage des poids entre les entrées du EC et le DG est de nature hebbienne :

$$\Delta W_{i,j}^{(EC,DG)} = \alpha^{(EC,DG)} DG_j (EC_i - W_{i,j}^{(EC,DG)}) \quad (4.14)$$

4.1.2.2 Apprentissage S-R

Les sorties apprises ont toujours un référentiel allocentrique. L'apprentissage est similaire à l'apprentissage procédural (section 4.1.1.2), mais les entrées sont ici les cellules de lieu (PC) et non les entrées visuelles.

Les réponses motrices de l'expert Local au lieu sont encodées par $N_{AC} = 36$ cellules d'actions (*Action Cells AC*). Chaque AC reçoit les entrées de toutes les PCs et code une direction de mouvement $\phi_i = \frac{2\pi i}{N_{AC}}$. L'activité des ACs représente la valeur (i.e., récompense espérée) d'un mouvement dans la direction correspondante et est calculée comme suit :

$$a_i^L(t) = \sum_{j=1}^{N_{PC}} r_j^{PC}(t) w_{ij}^L(t). \quad (4.15)$$

L'activité totale de la population est interprétée comme un code de population pour la direction Φ^L pour le prochain mouvement de l'animat, proposée par l'expert Local (Strösslin et al, 2005; Chavarriaga et al, 2005) :

$$\Phi^L(t) = \arctan\left(\frac{\sum_i a_i^L(t) \sin(\phi_i^L)}{\sum_i a_i^L(t) \cos(\phi_i^L)}\right). \quad (4.16)$$

Comme en 4.1.1.2, la mise à jour des poids synaptiques est réalisée selon l'équation suivante :

$$\Delta w_{ij}^L = \eta^L \delta^L(t) e_{ij}^L. \quad (4.17)$$

Ici, η^L est le taux d'apprentissage et $\delta^L(t)$ est l'erreur de prédiction de récompense :

$$\delta^L(t) = R(t+1) + \gamma^L A^L(\Phi^L(t+1)) - A^L(\Phi^L(t)), \quad (4.18)$$

Où $R(t)$ est la récompense délivrée au temps t , $0 < \gamma < 1$ le facteur d'atténuation de la prochaine récompense (Sutton et Barto, 1998), et a^L la Q-valeur de l'action exécutée au temps t , estimée par l'expert Local.

A l'instar de l'apprentissage procédural, la mise à jour est toujours réalisée, quel que soit l'expert qui a réalisé l'action. Les traces d'éligibilité e_{ij}^T accélèrent l'apprentissage en mémorisant les couples état-action expérimentés dans le passé :

$$e_{ij}^L(t+1) = r_j^{PC}(t)r_i^{AC}(t) + \lambda^L e_{ij}^L(t), \quad (4.19)$$

Où λ^L est le taux d'atténuation de la trace et r_i^{AC} correspond à :

$$r_i^{AC}(t) = \exp\left(\frac{\phi_i^L - \Phi^L(t)}{(2\sigma^L)^2}\right), \quad (4.20)$$

4.1.3 Un expert de lieu "déclaratif" : l'expert Planification

4.1.3.1 Entrées

Les cellules de lieu sont supposées ne pas encoder d'information de transition entre différents lieux. La représentation pour cet expert est construite sous la forme d'une carte topologique, i.e., un graphe. Les lieux et les actions pour rejoindre un lieu à partir d'un autre sont encodées dans les arcs reliant les deux nœuds associés. Le graphe n'inclut pas d'information métrique : bien que cela soit possible grâce aux cellules de grilles, il n'existe pas assez d'éléments permettant d'étayer l'existence de l'encodage d'une telle information chez le rat.

La carte topologique est une couche supplémentaire prenant en entrées les cellules de lieu et offrant en sortie une représentation éparse. D'un point de vue computationnel, cette représentation éparse offre l'avantage d'être peu coûteuse pour une planification ultérieure : en faisant baisser le nombre de nœuds, on baisse aussi le nombre de parcours que l'algorithme de planification aura à effectuer.

Le modèle que nous utilisons est une version simplifiée de celui de Martinet et al (2008), au sens où les mini-colonnes ne sont pas explicitement modélisées dans les colonnes et celles-ci ne sont constituées que de deux couches : la première est connectée aux PCs et renseigne, par son activation, sur la position de l'animat, la seconde permet la diffusion de l'information de but grâce aux connexions latérales entre les colonnes. Ainsi, les nœuds et liens entre eux sont appris indépendamment de la position du but, cette dernière étant apprise pendant la phase de planification.

Le graphe de planification offre un codage épars de l'activité des PC, comme les PCs offraient un codage épars de l'activité des ECs. Nous appliquons donc l'étape d'apprentissage du DG au graphe de planification. Ceci implique de spécifier une taille N^{PG} en terme de nombre de nœuds pour le graphe de planification. Ces nœuds sont ensuite reliés aux cellules du DG par

des poids $W^{DG,PG}$ initialisés aléatoirement (et remplissent ainsi automatiquement la condition qu'un nœud doit avoir au moins une connexion non nulle avec les PCs).

L'activation des nœuds r_j^P du graphe de planification est ensuite calculée comme suit :

$$r_j^P = f_j\left(\sum_i W_{i,j}^{(DG,PG)}, s^{PG}\right), \quad (4.21)$$

où f_j est la même fonction d'activation que celle utilisée pour calculer le taux d'activation des cellules du DG. L'apprentissage des poids $W^{DG,PG}$ est donc également hebbien :

$$\Delta W_{i,j}^{(DG,PG)} = \alpha^{(DG,PG)} r_j^P (DG_i - W_{i,j}^{(DG,PG)}) \quad (4.22)$$

L'apprentissage des liens entre nœuds est réalisé dans une étape à part (après création des nœuds) selon le même processus utilisé pour les PCs ad hoc. Cependant il est nécessaire de connaître les coordonnées de chaque nœud dans l'environnement afin de calculer l'orientation utilisée pour rejoindre un nœud de départ et un nœud d'arrivée. Cette étape est réalisée en parcourant tout l'environnement, et en stockant dans une matrice aux indices x et y le nœud le plus activé en cet endroit. Le centre de chaque nœud est ensuite calculé en moyennant ces coordonnées.

Une fois le graphe appris, les champs d'activation des nœuds codent aussi pour un lieu spécifique, comme le montre la figure 4.5 a. La taille des champs d'activation est plus importante que celles des PCs (figure 4.4 b) et codent donc pour un lieu de taille plus importante.

4.1.3.2 Apprentissage de recherche de graphe

Les sorties sont l'exécution d'enchaînement d'actions ayant préalablement été planifiées "mentalement". Elles ont toujours un référentiel allocentrique.

Le chemin optimal vers le but est déterminé par le mécanisme biologiquement plausible d'activation-diffusion (Burnod, 1991), basé sur l'algorithme de Dijkstra pour trouver un plus court chemin entre deux nœuds (Dijkstra, 1959). Il est implémenté comme suit :

Pendant la planification, l'expert doit d'abord déterminer son emplacement puis la direction vers le but (Figure 4.5 b). Deux valeurs sont donc associées à chaque nœud : une valeur de position et une valeur de but. La valeur de position correspond à l'activité r_i^P du nœud i . La valeur de but $G_i = 0$ lorsque la position du but est inconnue. Dans ce cas, l'expert propose une direction de mouvement aléatoire parmi les différentes actions possibles depuis le nœud courant. A l'inverse, lorsque l'emplacement du but est connu, la valeur de but du nœud le plus proche est définie à $G_* = 1$, et est propagée à tous les nœuds adjacents, atténuée par un facteur α^P . Ainsi cette valeur de but diminue au fur et à mesure que le nœud est éloigné, en terme de nombre de nœuds intermédiaires jusqu'au nœud But. La valeur de but G_i d'un nœud i à une distance de n nœuds du but est donc :

$$G_i = \alpha^n \quad (4.23)$$

Dans ce cas, la prochaine direction de mouvement Φ^i est donnée par le lien vers le nœud adjacent j de valeur de but la plus forte (Figure 4.5 b) :

$$\Phi^i = \Phi_j, j = \operatorname{argmax}_G \quad (4.24)$$

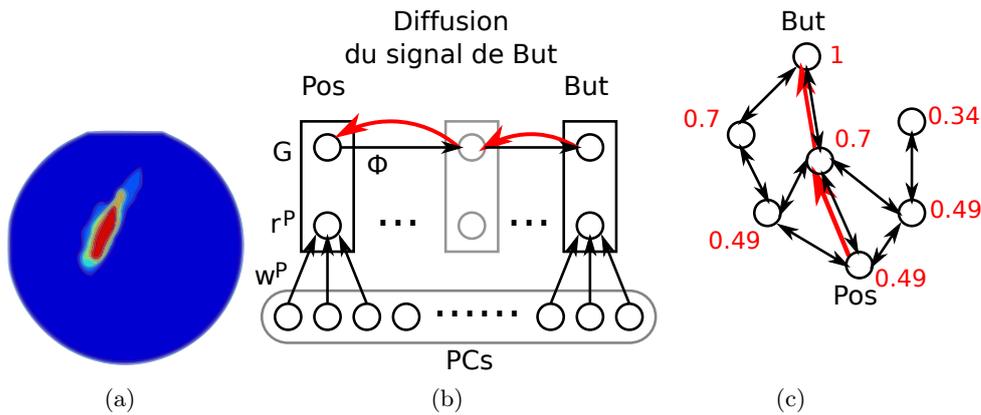


Figure 4.5 – (a) Champ d’activation d’un nœud appris à partir des PCs. (b) Structure du graphe de Planification : Les cellules de lieux sont en entrées du graphe, pondérées par les poids w_{ij}^P . La somme pondérée donne la valeur de position r^P . Lorsque le but est atteint, l’algorithme d’activation-diffusion assigne une valeur de but G_i à chaque nœud, amoindri par un facteur α . (c) L’animat (ici en Pos) rejoint le but en choisissant les nœuds adjacents qui maximisent la valeur de but (en rouge, trajectoire représentée par la flèche rouge).

4.1.4 Expert Exploration

Nous avons choisi de sélectionner un expert Exploration à part, qui puisse être sélectionné comme les autres experts (voir section 4.2). Cette décision découle notamment de l’hétérogénéité de la gestion de l’exploration dans les différents algorithmes d’apprentissage. Cela nous permet également d’évaluer le rôle de l’exploration à part entière dans la reproduction des expériences. D’autres raisons techniques seront évoquées sur le choix de cette exploration lorsque le réseau de sélection aura été présenté (section 4.2). L’exploration consiste en une action d’orientation choisie aléatoirement entre 0 et 2π . C’est le seul expert à ne pas dépendre d’entrées et à ne pas avoir d’apprentissage. Bien que peu biologiquement plausible dans sa forme actuelle, l’exploration permet de faciliter l’atteinte des lieux impossibles à visiter avec les autres experts.

4.1.5 Apprentissages réciproques des stratégies

Un mécanisme très important de notre modèle est la possibilité d’un expert d’apprendre du comportement généré par un autre expert. En effet, quel que soit l’expert de navigation qui a été choisi, lorsque l’animat atteint un but, le signal de récompense $R(t)$ est partagé entre tous les experts, indépendamment de l’expert qui a généré l’action effectuée. Les actions exécutées sous le contrôle, par exemple, de l’expert Planification permettent donc à l’expert Taxon d’ajuster ses poids en fonction. Réciproquement, si l’expert Taxon est responsable de l’action qui a mené au but, l’information de lieu de ce but sera également mise à jour par l’expert Planification. Ainsi, une coopération entre les stratégies peut être favorisée par cette implémentation.

4.2 Module de sélection des experts

Le critère de sélection du modèle de Chavarriaga et al (2005), dont notre modèle s'inspire, est limité par le fait qu'il ne peut pas sélectionner de stratégie n'utilisant pas d'apprentissage par renforcement, à l'image de la Planification présentée précédemment, ou de l'exploration aléatoire.

Le mécanisme de sélection présenté ci-dessous vise donc à satisfaire deux objectifs :

- Le modèle doit être adaptatif, c'est à dire qu'il doit apprendre selon les circonstances à choisir la stratégie la plus pertinente, sans a priori. Il doit donc s'adapter à un environnement potentiellement changeant, et, après une période d'apprentissage, sélectionner la meilleure stratégie selon l'environnement courant.
- Le critère de sélection ne doit pas reposer sur des valeurs inhérentes à la nature de l'apprentissage d'une stratégie. En effet, nous voulons pouvoir estimer la performance d'une stratégie, non en fonction d'une erreur de prédiction de récompense (experts procéduraux) ou d'une estimation de distance au but (graphe de planification), mais en fonction de son impact sur la performance générale de l'animat.

Notre réseau de sélection devra donc proposer à tout temps t , la stratégie qui permet de maximiser la probabilité d'atteindre le but, quelles que soient ses caractéristiques.

Nous retrouvons donc ici la formalisation habituelle d'un problème de décision, dont la résolution pourrait être apprise par renforcement. Dans cette optique, l'état courant serait représenté par l'activation actuelle des entrées des différents experts, et les actions possibles par les actions d'orientation de ces mêmes experts. Nous avons donc ici un premier bloc pour l'élaboration du réseau de sélection, à l'instar de celui de Chavarriaga et al (2005), à savoir un réseau de neurones connectant les entrées des experts à des actions traduisant la proposition d'orientation de chaque expert.

Toutefois, à l'inverse d'un TD-learning habituel, ces actions sont proposées par des experts, eux mêmes apprenant : cela signifie que les actions proposées ne sont pas figées dans le temps et que l'action résultant du choix d'un expert au temps t dans un état s pourrait être tout autre dans ce même état au temps $t + n$.

Le plus grand dénominateur commun entre plusieurs stratégies de navigation est la *proposition d'orientation* que fait chacune d'entre elles. Ainsi, si dans un même état, deux experts font une proposition d'orientation similaire qui mène au but, il s'avère cohérent de favoriser leur sélection la prochaine fois que cet état sera rencontré (même l'expert qui n'a pas été sélectionné). A l'inverse un expert dont l'action exécutée provoque une collision avec un obstacle sera sanctionné, de même que les experts proposant des actions proches, tandis que ceux proposant des actions opposées seront favorisés.

4.2.1 Critère de sélection

Pendant l'apprentissage, le modèle doit sélectionner parmi les experts proposant une orientation, la prochaine direction de mouvement. Le réseau de sélection décrit ici est conçu pour estimer la performance de chaque expert et choisir le meilleur à chaque pas de temps (Chavarriaga et al, 2005). La structure du réseau est un réseau de neurones à taux *feed-forward*, à une couche, liant les entrées des n experts (E_1 à E_n), chacun normalisés de sorte à ce que les moyennes des entrées soient égales entre tous les experts. Dans le modèle, le réseau est constitué de n unités $k \in \{E_1 \dots E_n\}$, chacune correspondant à un expert séparé. L'activité d'une unité k est appelée "valeur de sélection" de l'expert correspondant. Les entrées de ces unités sont fournies par les activités des entrées des experts correspondants (Figure 4.6). Les valeurs de sélection g^k sont calculées comme suit :

$$g^k(t) = \sum_{i=1}^n z_{E_i}^k \quad (4.25)$$

où, pour chaque expert E_i :

$$z_{E_i}^k = \sum_{j=1}^{N_{E_i}} z_j^k(t) r_j(t) \quad (4.26)$$

où N_{E_i} est le nombre d'entrées de l'expert E_i , z_j^k est le poids de connexion entre l'unité k du réseau et l'entrée j .

Chaque expert propose à chaque pas de temps des directions de mouvement candidates Φ_k . Les valeurs de sélection sont comparées pour choisir le prochain mouvement Φ^* en utilisant un schéma winner-takes-all :

$$\phi^k(t); k = \operatorname{argmax}_i(g^i(t)) \quad (4.27)$$

Les poids de connexion pour les valeurs de sélection des experts E_i sont mis à jour par un algorithme Q-learning. La règle de mise à jour dans ce cas est :

$$\Delta z_j^k = \xi^G \delta^G(t) e_j^k(t+1), \quad (4.28)$$

où ξ^G est le taux d'apprentissage du réseau de sélection, et $\delta^G(t)$ l'erreur de prédiction de récompense :

$$\delta^G(t) = R(t+1) + \gamma^G \max_k (g^k(t+1)) - g^{k^*}(t), \quad (4.29)$$

$R(t)$ étant la récompense délivrée au temps t , γ^G le facteur d'atténuation de la récompense du réseau de sélection et g^{k^*} la valeur de sélection de l'expert choisi au temps t .

Comme pour l'expert Taxon, les traces d'éligibilité e_j^k de l'expert k permettent au réseau de renforcer les experts sélectionnés dans le passé :

$$e_j^k(t+1) = \Psi(\Phi^*(t) - \Phi^k(t)) r_j^k(t) + \lambda^G e_j^k(t), \quad (4.30)$$

où λ^G est le facteur d'atténuation de la trace. Le terme $\Psi(\Phi^*(t) - \Phi^k(t))$ peut être considéré comme la version discrète de la généralisation des actions de l'expert Taxon, où :

$$\Psi(x) = \exp(-x^2) - \exp(-\pi/2). \quad (4.31)$$

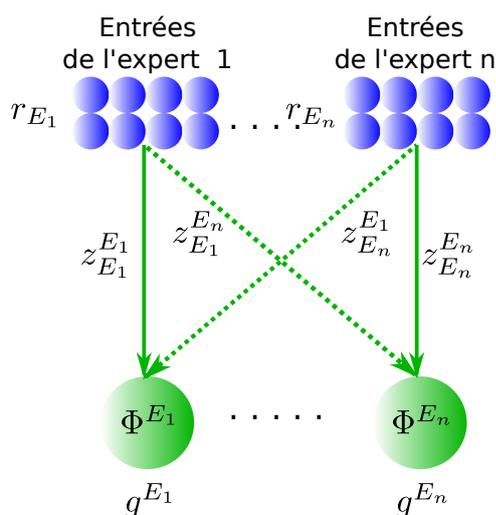


Figure 4.6 – Réseau de sélection. Les entrées des Experts E_k sont reliées aux unités du réseau de sélection. Les valeurs de sélections g^{E_k} sont les sommes pondérées des entrées r_j par les poids de connexion $z_j^{E_k}$. Un des experts est sélectionné selon un schéma *winner-takes-all*.

Ce terme, pouvant être vu comme la version discrète du profil de généralisation (Strösslin et al, 2005), garantit que le renforcement d'un expert sera d'autant plus important que sa direction de mouvement proposée sera proche de celle sélectionnée. À l'inverse deux experts proposant deux directions opposées auront également des renforcements opposés. Le réseau de sélection et sa mise à jour sont illustrés Figure 4.6.

La sélection entre experts est effectuée à chaque pas de temps, à moins que l'expert Exploration ne soit choisi, auquel cas la direction sera maintenue pendant trois pas de temps, afin d'éviter les cas d'oscillations comportementales où l'animat ne parviendrait pas à sortir d'une zone donnée.

4.2.2 Modularité de la sélection

La sortie du réseau de sélection étant une orientation, il est alors possible de créer des "méta-experts" constitués de plusieurs experts sélectionnés par un réseau de sélection, ce méta-expert sera à son tour connecté à un réseau de sélection : les entrées de chaque expert du méta-expert sont connectées au réseau de sélection, et la direction de mouvement choisie par le méta-expert est ensuite "remise en jeu" par le réseau de sélection principal. Aucune autre information que ces deux-là n'est donc nécessaire au réseau de sélection pour mettre en œuvre la sélection de stratégies de navigation. Cela permet de simuler notamment des stratégies de guidage complexe en présence de plusieurs amers, un méta-expert de guidage étant représenté par des experts de guidage ancré chacun sur un amer. Il n'existe pas de limite théorique au nombre d'experts pouvant être sélectionnés.

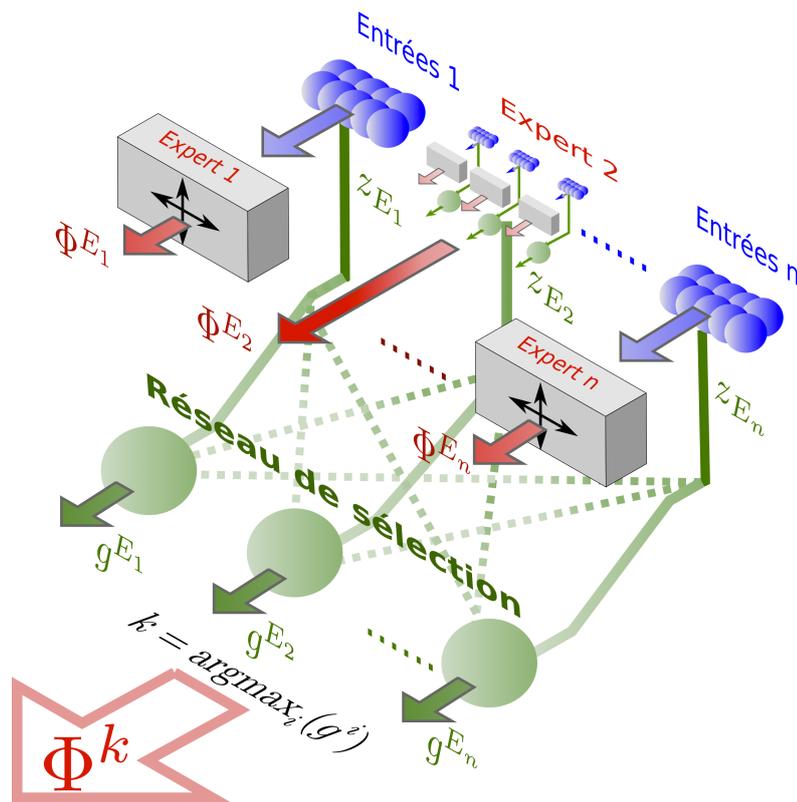


Figure 4.7 – Vue générale du modèle. Chaque expert prend des entrées (cellules bleues) pour apprendre sa stratégie, ces entrées sont également reliées au réseau de sélection (liens z_{E_i} en vert) qui calcule la valeur de sélection de ces experts (cellules vertes), en fonction des orientations Φ proposés par les experts (flèches rouges). L'expert 2 est un méta-expert constitué de plusieurs experts sélectionnés par un sous-réseau (par exemple un méta-expert de guidage qui choisit successivement des indices visuels parmi plusieurs).

En résumé, la structure générale du modèle, pour n experts, est illustrée dans la Figure 4.7.

4.3 Analyse des données

Groupes d'experts. Selon les tâches à effectuer, seuls certains groupes d'experts suivants seront choisis a priori. Les groupes "contrôles" auront accès à la fois aux stratégies de guidage et de lieu, mais les apprentissages de ces dernières peuvent différer selon les groupes (Taxon ou Direction pour le guidage, Locale ou Planification pour les stratégies de lieu). Les groupes lésés n'auront accès à l'inverse qu'à un seul type de stratégie (de guidage ou de lieu, mais pas les deux) car nous supposons, de façon sans doute trop dichotomique par rapport à la littérature biologique, que les animaux ayant une inactivation d'une structure impliquant telle ou telle stratégie ne pourra pas du tout l'exprimer (i.e., es animaux hippocampiques ne pourront

exprimer qu'une stratégie de guidage ; ceux lésés du DLS qu'une stratégie de lieu). Ainsi la lésion est réalisée dans notre modèle en "désactivant" la sélection de la stratégie correspondante : celle-ci est alors ignorée, ses entrées ne sont pas prises en compte dans le réseau de sélection, de même que l'unité de sélection correspondante.

Les groupes seront alors abrégés comme suit (voir leur code couleur Figure 4.8) :

Groupes "contrôles"

- Taxon-Planification : groupe TP
- Direction-Planification : groupe DP
- Taxon-Local : groupe TL
- Direction-Local : groupe DL

Groupes "lésés"

- Taxon (équivalents animaux hippocampiques) : groupe T
- Direction (équivalents animaux hippocampiques) : groupe D
- Planification (équivalents animaux lésés du DLS) : groupe P
- Local (équivalents animaux lésés du DLS) : groupe L

Notons que chacun de ces groupes sera toujours constitué d'un "expert" Exploration proposant des directions aléatoires.

Les entrées seront soit séparées (indices proximaux pour experts Taxon ou Direction, indices distaux pour expert Local ou Planification), soit partagées entre tous les experts. Les symboles des groupes seront alors associés avec **"-sep"** en cas d'entrées séparées, et **"-par"** en cas d'entrées partagées.

Dans le cas d'entrées séparées, nous laissons le modèle choisir la meilleure stratégie, même si l'un des experts ne reçoit qu'une entrée bruitée. Dans le cas de la seule présence d'amers distaux, les experts Taxon et Direction ne pourront pas construire de carte de navigation et se comporteront comme l'expert Exploration. Dans le cas de la seule présence d'amers proximaux, l'expert Planification pourra quand même construire une carte avec ses seules entrées odométriques.

Ainsi, nous pourrions évaluer la pertinence de deux hypothèses : la séparation en amont des entrées (entre amers distaux et proximaux) ou leurs sélection a posteriori, sur la base des performances des stratégies.

L'analyse des données est commune à tous les dispositifs. Le nombre d'animats est de 50 par expériences. Celles-ci seront menées avec les entrées séparées par experts, ou les entrées partagées entre experts. Nous entendons par entrées séparées le fonctionnement du modèle avec des entrées sensorielles bien départagées entre experts : l'expert Taxon, ou Direction, dans le cas de plusieurs amers, ne prend en entrée que des amers proximaux tandis que l'expert Planification ne construit sa carte qu'avec des amers distaux. Dans le cas des entrées partagées, l'ensemble des experts partage les mêmes données sensorielles : l'expert Taxon, ou Direction, reçoit les représentations à la fois des amers proximaux et distaux, tandis que l'expert Planification construit sa carte

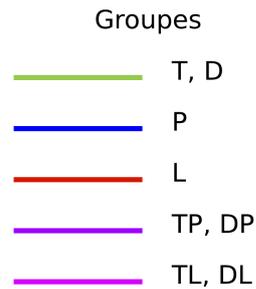


Figure 4.8 – Codes couleurs des graphes représentant les différents groupes

avec les deux types d'amers.

Nous décrivons ici les principales méthodes utilisées pour analyser les performances et comportements des animats.

Estimation des performances. Les performances des groupes Contrôle, Taxon, Locale et Planification sont statistiquement vérifiées par la comparaison de leurs temps d'échappement moyens (i.e., le nombre de pas de temps avant de rejoindre le but). Le test de Wilcoxon est utilisé pour les échantillons provenant du même groupe tandis que la comparaison d'échantillons de groupes différents est vérifiée par un test de Mann Whitney. L'apprentissage des groupes est représenté par l'évolution des temps d'échappement (en pas de temps) au cours des sessions. Les différences significatives entre les apprentissages des groupes sont indiquées par une ($p < 0.05$), deux ($p < 0.01$) ou trois ($p < 0.001$) étoiles.

Calcul des taux de sélection et d'occupation des experts. Le comportement des animats est analysé par le *taux de sélection par essai* d'un expert : il correspond au nombre de fois où cet expert est choisi sur la longueur totale de la trajectoire, pondéré sur 100 expériences.

Pour certaines expériences, nous aurons besoin de connaître le comportement de l'animat dans une zone donnée, nous utiliserons alors le *taux d'occupation du but* est défini par le nombre de fois qu'un animat se rend à proximité du but, divisé par la longueur totale de la trajectoire.

Le *taux de sélection au but* d'un expert correspond au nombre de fois où cet expert est sélectionné dans une zone autour du but. La taille de cette zone dépend de l'expérience simulée : il pourra s'agir du quadrant du but (un quart de disque centré autour du but), d'un octant (huitième de la surface) ou d'une surface de 0.4 m^2 autour du but.

Trajectoires typiques. Des trajectoires typiques d'un individu peuvent être tracées en mentionnant par le code couleur l'expert qui est sélectionné à chaque pas de temps.

Cartes de navigation L'apprentissage d'un expert de guidage ou Local (procédural) d'un individu peut être illustrée par sa carte de navigation : il s'agit d'un champ de vecteurs représentant, en chaque point de l'environnement, la direction préférée par l'expert. La carte topologique d'un expert Planification (déclaratif) peut être également représentée sous forme d'un graphe orienté, et est calculée en moyennant les coordonnées des nœuds ayant une activation maximale lors d'un

passage complet sur l'environnement. Les flèches reliant les nœuds représentent les orientations choisies pour rejoindre le but courant après planification.

La plupart des simulations sont validées sur la reproduction de certains résultats observés dans les expériences animales. Les Figures concernant le groupe qui montre les résultats les plus proches des résultats originaux sont alors entourées.

Quatrième partie

Résultats : Simulations de l'influence
d'amers proximaux et/ou distaux sur
l'apprentissage spatial

Chapitre 5

Présence d'un seul type d'amer

Ce chapitre est destiné à estimer comment les experts de guidage et de lieu du modèle interagissent lorsque dans l'environnement il n'y a qu'un seul type d'indice, proximal ou distal. Des reproductions d'expériences qui suivent ces conditions et qui mandatent les théories associative et cognitive seront analysées en termes de ces interactions.

Sauf quelques exceptions, l'environnement simulé décrit ici est semblable pour toutes les expériences de ce chapitre : la piscine de Morris simulée a 2 mètres de diamètre (dimension la plus communément utilisée dans les expériences réelles). Quatre positions de départ sont définies, correspondant aux quatre principaux points cardinaux (Figure 5.1). Le choix est réalisé afin d'éviter deux mêmes positions de départ consécutives. Lors d'un essai, l'animat a 350 pas de temps au maximum pour trouver l'emplacement de la plateforme, au-delà desquels il est automatiquement guidé vers elle. En tout, 8 emplacements possibles ont été définis pour le but (Figure 5.1). Un emplacement est pseudo-aléatoirement choisi au début de l'expérience et reste stable durant toute l'expérience.

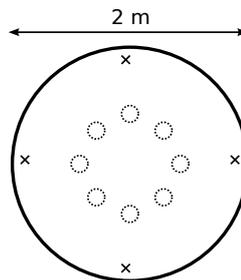


Figure 5.1 – L'environnement simulé : une piscine de Morris de 2m de diamètre. Croix : points de départ de l'animat, cercles pointillés : emplacements possibles du but.

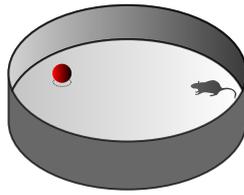


Figure 5.2 – EXP1 : L'environnement simulé : piscine de Morris avec but fixe et balise.

5.1 Présence des seuls amers proximaux

En présence d'un seul amer proximal - but visible ou indiqué par un amer situé dans la piscine - ou de deux amers signalant le but, l'hypothèse est que les experts non dépendants de l'hippocampe prennent le contrôle, si les essais sont suffisamment nombreux (cf. Etat de l'art, chapitre 2).

Nous simulons les conditions suivantes, où les amers distaux sont totalement absents (deux protocoles que nous avons conçus et un protocole issu d'une expérience animale) :

- Un environnement avec une balise à l'endroit du but (EXP1).
- Un environnement avec un amer à côté d'un but invisible (EXP2).
- Un environnement expérimental où deux amers se trouvent à côté d'un but invisible (McGregor et al, 2004a, expérience 2).

La dernière condition fait partie d'une expérience qui adresse la nature associative de l'apprentissage spatial et que nous allons poursuivre dans le Chapitre suivant.

5.1.1 But visible fixe ou balisé (EXP1)

De nombreuses expériences montrent que les stratégies de guidage sont supposées être privilégiées dans cette condition (cf Etat de l'art, Chapitre 6), la pauvreté de l'environnement ne permettant pas la création d'une carte topologique opérationnelle. Une stratégie Taxon égocentrée est souvent évoquée, mais la stratégie Direction peut aussi être efficace, facilitée par la nature allocentrée de son cadre de référence.

Six groupes vont être simulés : Taxon-Planification et Direction-Planification, experts aux entrées séparées et partagées (TP-sep, TP-par, DP-sep, DP-par, de même que T et D). Nous n'intégrerons pas d'expert Local dans cette section, pour n'estimer que l'effet de l'interaction avec un expert ayant un autre mode d'apprentissage que les deux experts de guidage.

5.1.1.1 Protocole et résultats simulés

L'environnement utilisé est celui décrit Figure 5.2. La balise a un diamètre de 13 cm, soit le même diamètre que le but. Celui-ci est situé à l'un des 8 emplacements possibles et reste fixe. Une expérience est composée de 10 sessions de 4 essais.

Tous les groupes apprennent la tâche, le groupe T est le plus rapide. Tous les groupes montrent un apprentissage rapide (Figure 5.3 a, c), puisque les temps d'échappement

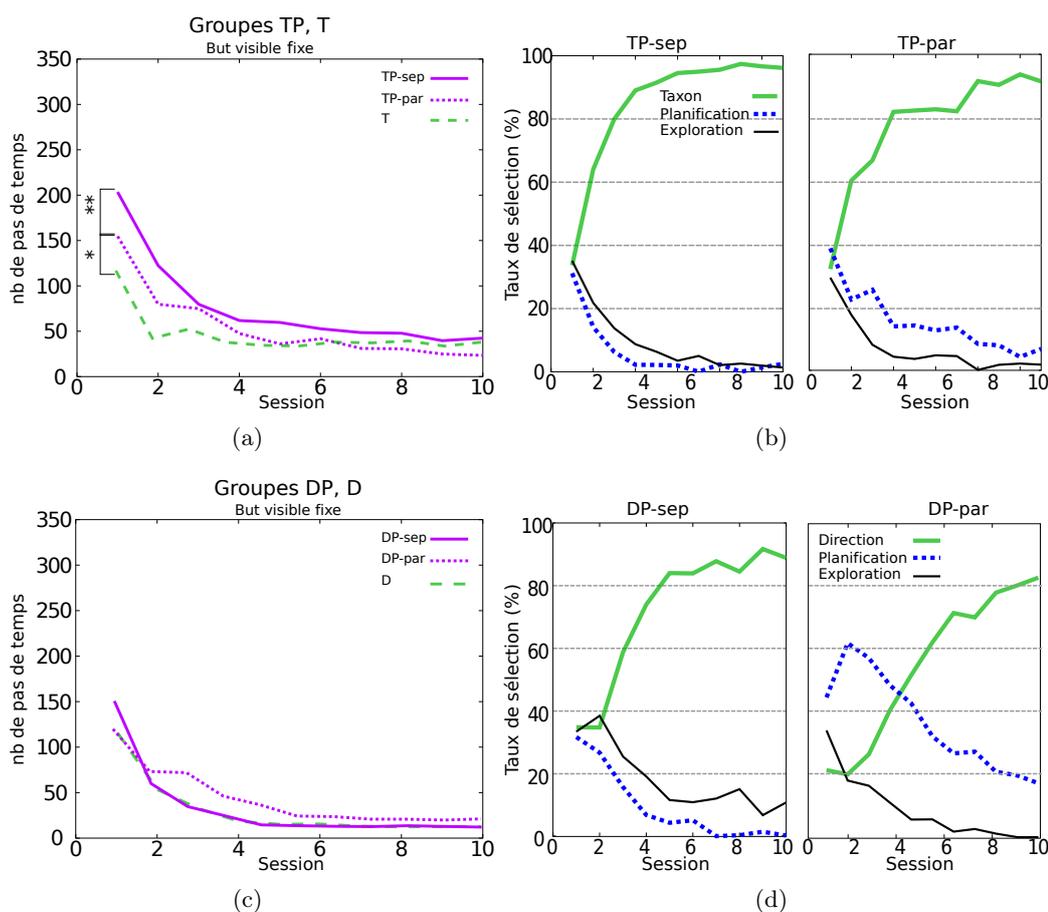


Figure 5.3 – **But visible ou balisé fixe.** Groupes TP-sep, TP-par, T : (a) Temps d'échappement de l'animat selon les entrées ; (b) Taux de sélection par session, selon les entrées et la nature de la stratégie de guidage ; (c - d) idem avec Groupes DP-sep, DP-par, D.

en fin d'expérience sont autour de 50 à 25 pas de temps. Le groupe T montre les meilleures performances au début de l'apprentissage (45 pas de temps dès la deuxième session, contre 55 pour le groupe D). L'expert Direction contribue à une meilleure efficacité des performances en fin d'expérience. Comme nous l'avons supposé plus haut, la nature allocentrée de cet expert semble accélérer la rapidité d'atteinte du but.

La stratégie de guidage prend toujours le contrôle. Les taux de sélection des stratégies (Figure 5.3 b, d) montrent que les experts Taxon ou Direction sont dans tous les cas très majoritairement sélectionnés. Lorsque l'expert Planification a accès à l'amer proximal (Figure 5.3 b, d, droite), sa sélection augmente ($p < 0.001$). L'expert Direction est alors significativement moins sélectionné que dans les autres conditions (sur les dix derniers essais : 77% de sélection contre 91 %, 96% et 94% avec $p < 0.001$ pour les trois autres groupes) : la nature allocentrée des deux experts est en effet redondante.

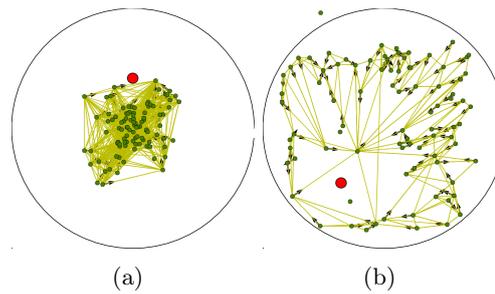


Figure 5.4 – **But visible ou balisé fixe (en rouge) (EXP1)**. Cartes de l'expert Planification construites (a) dans les groupes à entrées séparées (entrées seulement odométriques), (b) dans les groupes à entrées partagées (avec un seul amer proximal : le but)

L'expert Planification peut construire une carte sur la base d'un seul amer proximal. Une carte de planification est effectivement créée. A partir des seules informations odométriques (entrées séparées), elle ne contient aucune information topologique (Figure 5.4 a). Cet expert propose alors des orientations semblables à celles de l'exploration. Sur la base de ces informations et de la présence d'un seul amer proximal (entrées partagées, Figure 5.4 b), l'ensemble de ses nœuds est regroupé sur un cercle autour de l'amer et la carte est très éparse. On constate, particulièrement pour les groupes constitués de l'expert Taxon, qu'elle peut diminuer les performances.

5.1.1.2 Discussion partielle

Comme prévu, les résultats confirment que les experts de guidage sont efficaces pour atteindre un but fixe visible ou directement indiqué. Ces résultats vont dans le sens des hypothèses émises par les auteurs d'expériences similaires, notant le cadre de référence égocentré de la stratégie. On note que l'expert Direction peut améliorer les performances, après un entraînement plus long.

Contrairement aux hypothèses de départ, l'expert Planification peut participer à l'apprentissage d'un but fixe, même en l'absence d'amers distaux, car une carte très limitée peut se construire à partir d'entrées odométriques et/ou d'un but visible. Dans le premier cas, l'expert n'apportera que de l'exploration. Cependant il n'a pas tendance à améliorer l'apprentissage de l'expert Taxon égocentré, du moins en début d'expérience. La sélection de l'exploration retarde celle de l'expert Taxon.

5.1.2 Un seul amer à côté d'un but invisible (EXP2)

En mettant à profit les résultats précédents, l'hypothèse est que les experts de guidage prennent le contrôle, avec toutefois une participation de l'expert de lieu. On s'attend cependant à ce que l'expert Taxon soit moins performant que précédemment, puisqu'il doit chercher "à l'aveugle" la plateforme invisible autour de l'amer. L'expert Direction devrait être plus efficace que l'expert Taxon, puisqu'il apprendra à s'orienter dans une direction particulière par rapport

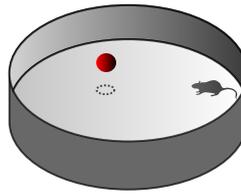


Figure 5.5 – EXP2 : L'environnement simulé : piscine de Morris avec but fixe et amer à côté du but.

à l'amer.

5.1.2.1 Protocole et résultats simulés

Les groupes et le protocole sont les mêmes que précédemment, à l'exception du but qui est caché, et la présence d'un amer indiquant sa présence situé 30 cm au nord du but (Figure 5.5).

Pour tous les groupes, l'apprentissage des experts de guidage est moins performant que précédemment (but visible). On s'attendait à cette chute de performance, tant au niveau de la performance générale que de l'amélioration de l'apprentissage au cours du temps (Figure 5.6 a, c). Les experts de guidage (groupes T et D), seuls ou accompagnés d'un expert de lieu avec carte "odométrique" (TP-sep et DP-sep) sont les moins performants. Les groupes avec Taxon ont un apprentissage extrêmement réduit (de 250 à 200 pas de temps), mais significatif ($p < 0.01$ sur les quatre premières sessions).

L'expert de lieu coopère avec l'expert de guidage pour améliorer significativement les performances : le groupe DP-par est le plus performant. L'accès à l'amer pour l'expert Planification améliore en effet les résultats des groupes TP-par et plus encore pour le groupe DP-par (respectivement $p < 0.001$, 75 pas de temps à la fin des 10 sessions au lieu de 230, Figure 5.6 a ; et $p < 0.001$, fin des 10 sessions, 40 pas de temps au lieu de 96, $p < 0.001$, Figure 5.6 a, c). Cela indique une coopération entre les deux experts pour atteindre la plateforme dès le début de l'expérience, les taux de sélection de l'expert Planification étant plus importants que lorsque l'amer est situé sur le but (Figure 5.6 b, d, droit). Les deux experts de guidage prennent néanmoins le contrôle - plus rapidement pour l'expert Taxon.

5.1.2.2 Discussion partielle

On notera que dans le cas d'un amer situé à côté du but, les moins bonnes performances (par rapport à celles de la recherche d'un but visible) sont dues essentiellement aux experts de guidage et non à l'intervention de l'expert de Planification, qui augmente une capacité d'exploration ou d'orientation vers la position de la plateforme.

Ces résultats montrent que l'utilisation d'une stratégie de lieu dans des conditions environnementales très pauvres ne peut en effet être écartée et peut même intervenir majoritairement en

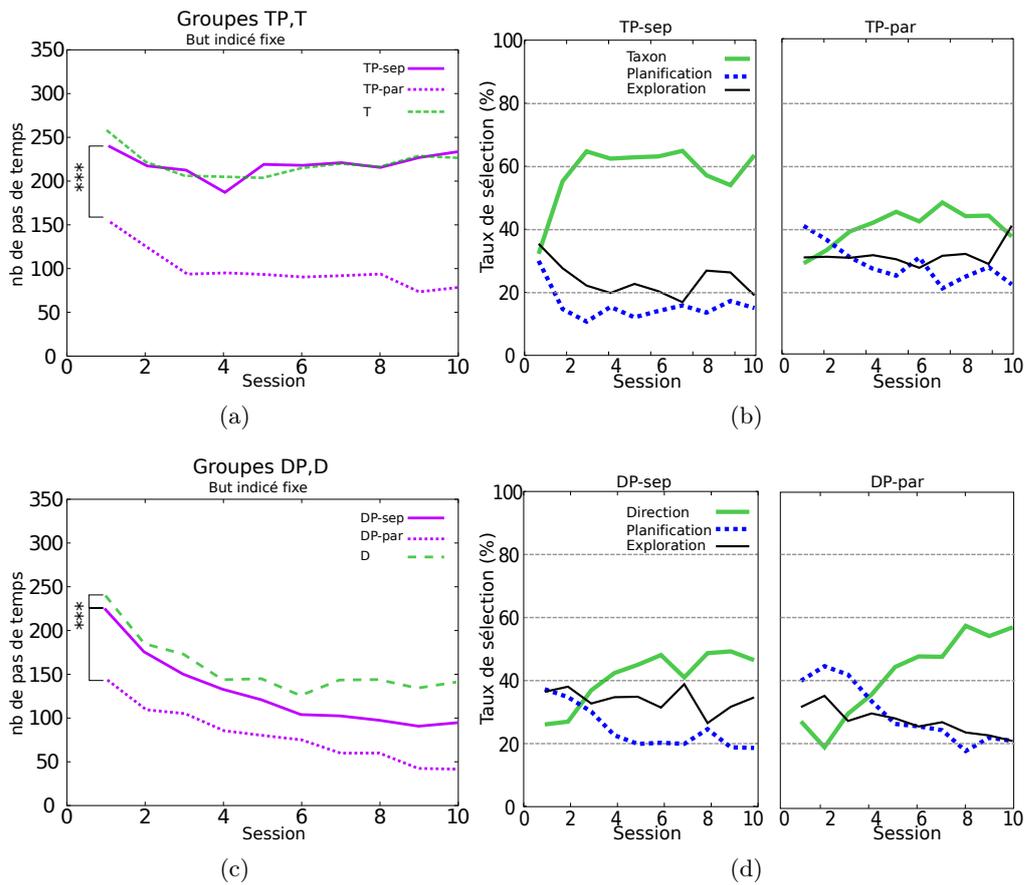


Figure 5.6 – **Un seul amer proximal à coté du but (EXP2)** Groupes TP-sep, TP-par et T : a) Performances de l'animat selon les entrées. b) Taux de sélection par session selon les entrées et la nature de la stratégie de guidage. c-d) idem pour les groupes DP-sep, DP-par et D.

début d'expérience. Elle pourrait être confirmée par la comparaison d'animaux hippocampiques (équivalents des groupes T, D, TP-sep, DP-sep) et d'animaux contrôles (équivalents des groupes TP-par ou DP-par), dont les performances devraient être significativement différentes. Pour opter pour un référentiel égocentré ou allocentré de la stratégie de guidage des animaux, un indice pourrait être de constater que s'ils ne progressent que très peu au cours des sessions, alors leur référentiel pourrait être égocentré.

5.1.3 Présence de deux amers à côté du but (McGregor et al, 2004a)

Cette expérience avait pour but de vérifier si un apprentissage simultané avec des amers proximaux et distaux pouvait ou non occulter l'apprentissage des amers proximaux. Elle vérifiait aussi si la position de la plateforme pouvait être mieux apprise à l'aide de deux amers et non d'un seul. Nous n'allons tester dans ce paragraphe que cette dernière hypothèse (puisque cette section concerne uniquement des situations où les amers distaux sont absents), mais nous reprendrons la suite de cette expérience dans la section Interactions d'amers (p. 130), où un autre groupe est testé en présence de ces amers proximaux, mais aussi d'amers distaux. Nous reproduisons ici le protocole du groupe "Rotate" de l'expérience 2 de McGregor et al (2004a). Pour ces animaux testés dans une piscine de Morris, deux amers proximaux et la plateforme sont déplacés (par rotation de 90°, d'où le nom du groupe) dans l'expérience 1 à chaque essai pour que les amers distaux ne soient pas pris en compte dans l'apprentissage. Dans l'expérience 2, des rideaux occultent la pièce, donc le groupe Rotate n'a pas du tout accès aux amers distaux.

Notre hypothèse est que, compte tenu des résultats précédents, les deux experts de guidage égocentré et allocentré peuvent contribuer à la tâche. Les experts de guidage et de lieu vont coopérer et utiliser les deux amers et non un seul, car la localisation de la plateforme sera plus précise.

5.1.3.1 Protocole et résultats observés

Le groupe Rotate doit rejoindre dans une piscine de Morris (2 m de diamètre) une plateforme cachée (10 cm), indiquée par deux amers proximaux. Des rideaux noirs occultent la pièce. Les deux amers A et B (respectivement 8 et 12 cm) ont été choisis et placés de sorte qu'aucun ne permette plus facilement de rejoindre le but : ils sont placés à 30 cm de façon à former un triangle équilatéral avec lui. L'entraînement se déroule sur 15 sessions de 4 essais (Figure 5.7 a, gauche). Les essais 2 et 4 des sessions 12 et 14 sont des essais de test. Dans la session 12, un seul amer (A ou B, selon les animaux) est disponible (Figure 5.7 a, milieu) et les deux le sont dans la session 14 (Figure 5.7 a, droite), ceci afin de vérifier si les deux amers sont pris en considération. La plateforme reste au même endroit dans l'essai 2 (noté "correct") et est déplacée à l'opposé du segment constitué par les amers dans l'essai 4 (noté "opposé"), ceci afin de vérifier si la plateforme est trouvée par exploration aléatoire ou non. L'ordre de ces essais (essai 2 correct/opposé, essai 4 opposé/correct) est variable selon les individus.

La courbe d'apprentissage des animaux est similaire à celle que les auteurs avait mesurée dans l'expérience 1 (Figure 5.7 b). Les auteurs ont comparé le temps nécessaire aux animaux pour rejoindre la plateforme dans chacun des essais test. Dans la session 12 (un seul amer présent),

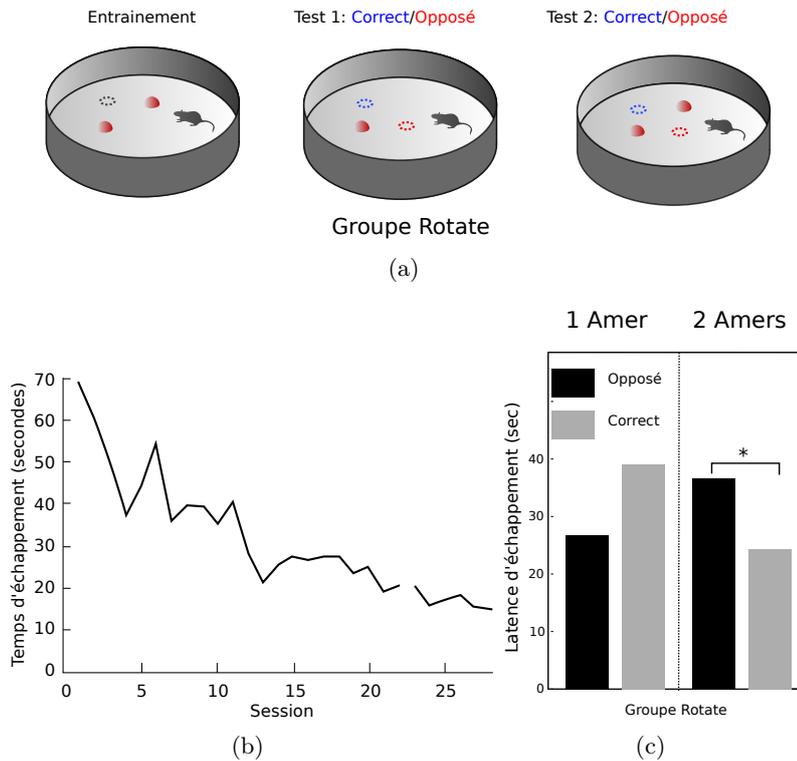


Figure 5.7 – **Expérience de McGregor et al (2004a)**. (a) Protocole original, voir texte pour détails. (b) Apprentissage moyen des animaux du groupe Rotate de l'expérience 1, qui est similaire à l'expérience 2 simulée ici, à part le nombre de sessions (12 seulement dans l'expérience 2). (c) Résultats aux essais de test : les animaux du groupe Rotate présentent une différence significative de temps d'échappement pour atteindre la localisation "correcte" ou "opposée" de la plateforme, seulement dans le cas de la présence de deux amers.

les animaux du groupe Rotate mettent des temps non significativement différents pour rejoindre la plateforme dans les essais correct et opposé (Figure 5.7 c, gauche). Dans la session 14 (deux amers présents), les animaux sont significativement plus performants dans l'essai correct que dans l'essai opposé (Figure 5.7 c, droite). L'apprentissage préalable effectué avec deux amers n'est donc pas transféré à un seul. Selon la théorie associative, ceci pourrait s'expliquer par une valeur de prédiction du but partagée entre les deux amers.

5.1.3.2 Protocole et résultats simulés

L'environnement ne contient donc aucun amer distal. La piscine et les deux amers proximaux A et B sont de même taille que dans l'expérience originale avec la même configuration. Les groupes simulés sont les groupes TP, T, DP et D (entrées séparées et partagées). La reproduction de cette tâche, avec la présence de deux amers A et B, n'est possible avec notre modèle que s'il considère un expert de guidage propre à chaque amer. Dans les groupes TP et DP, les experts de guidage (Taxon_A , Taxon_B ou Direction_A , Direction_B) sont pre-sélectionnés par un sous-module de sélection et l'orientation gagnante est confrontée à celles de l'expert de lieu et de l'exploration. Les deux amers proximaux peuvent donc être à tout moment en concurrence. Dans le cas des entrées séparées, l'expert Planification construit une carte uniquement par odométrie. Il a donc une fonction plutôt exploratoire. Dans le cas des entrées partagées, cet expert a accès aux deux amers pour bâtir sa carte.

La comparaison des résultats à l'entraînement et aux essais tests de tous les groupes avec les résultats observés montre que seuls les groupes avec expert Direction sont globalement conformes au comportement des animaux (Figure 5.8 a-d).

Les groupes D et DP sont les plus performants. Pendant l'entraînement, l'apprentissage des groupes TP-sep est très limité (de 295 à 200 pas de temps, Figure 5.8 a). Lorsque l'expert de lieu utilise les deux amers, il améliore significativement des performances (TP-sep/TP-par : $p < 0.001$), – similaires au groupe T. L'apprentissage du groupe DP est nettement meilleur (Figure 5.8 b). Lorsque les deux experts ont accès aux deux amers (DP-par), les performances sont également significativement améliorées ($p < 0.01$, 50 pas de temps à la fin de l'apprentissage contre 85 pour le groupe à entrées séparées) – et similaires au groupe D.

Contrairement aux résultats concernant un amer situé à côté du but, les experts de guidage combinant les deux amers sont donc aussi performants que lorsqu'ils sont associés à un expert de lieu construisant sa carte sur les deux amers. Le rôle de la planification est essentiellement exploratoire : on remarque en effet que la planification, présente dans les groupes TP et DP, est remplacée par une sélection importante d'exploration aléatoire, tout aussi efficace, dans les groupes T ou D (Figure 5.9 Taxon : a ; Direction : c)

Contrairement à l'expert Direction, l'expert Taxon occulte l'un des amers. Pendant l'entraînement, les taux de sélection des experts montrent que la prise de contrôle des deux experts de guidage s'effectue très différemment. Alors que l'expert T_A occulte immédiatement l'amer B (Figure 5.9 a), les deux experts D_A et D_B sont conjointement sélectionnés pendant tout l'entraînement (Figure 5.9 c). L'occultation de l'un des amers s'explique par le fait que les

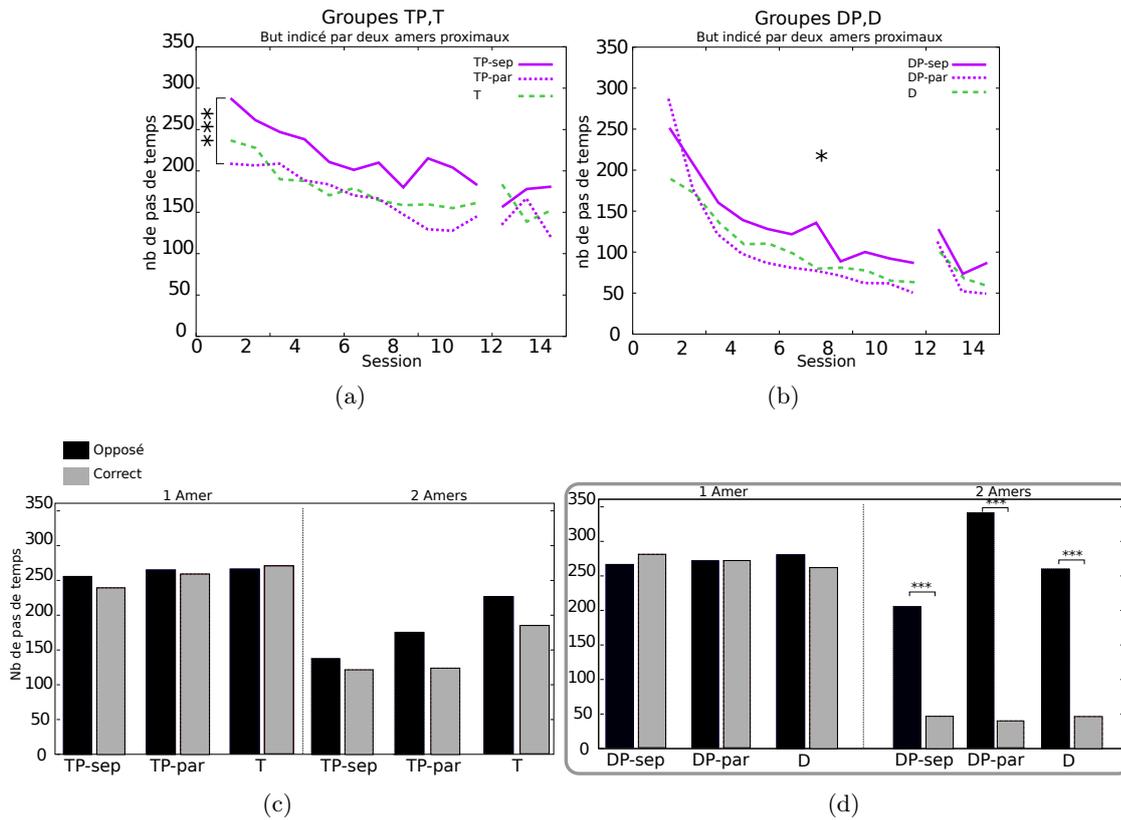
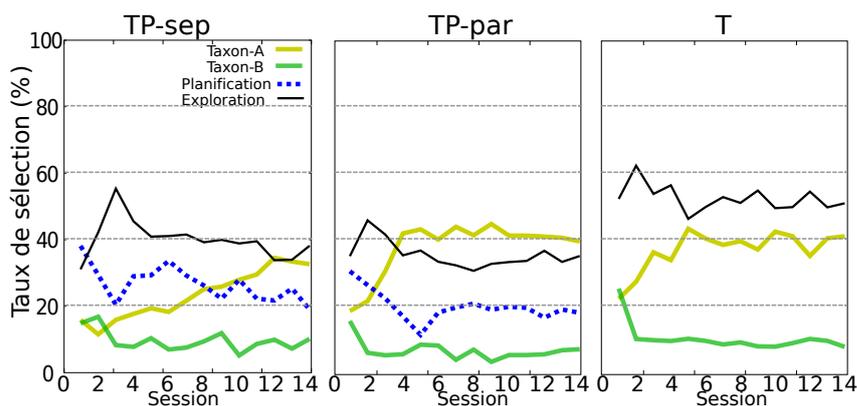
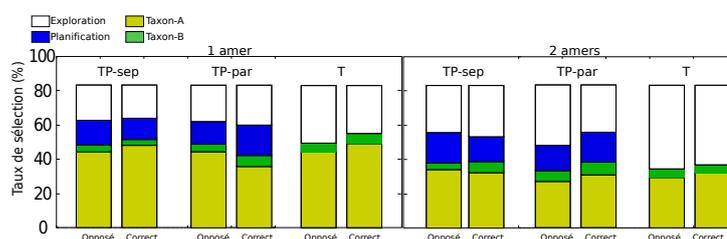


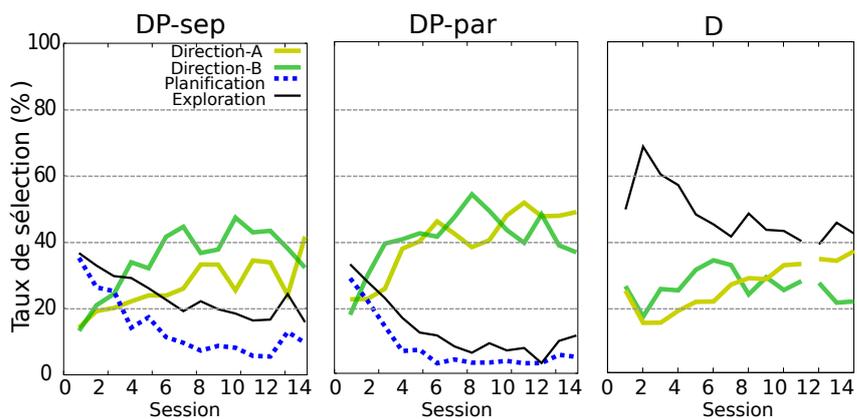
Figure 5.8 – **Simulation de l'expérience 2 de McGregor et al (2004a)** (a-b) Groupes TP-sep et -par, DP-sep et -par, groupes T et D : performances des apprentissages selon les entrées ; (c-d) Essais test pour ces groupes : taux d'échappement avant d'atteindre les localisations "opposée" et "correcte" de la plateforme, en présence d'un ou deux amers.



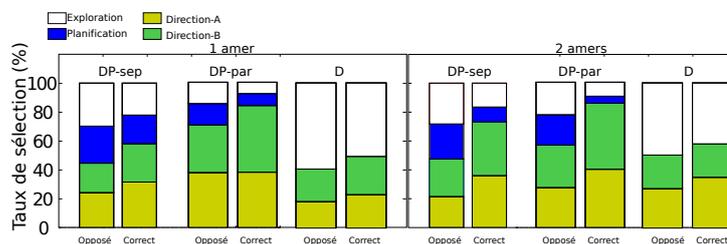
(a)



(b)



(c)



(d)

Figure 5.9 – Simulation de l'expérience 2 de McGregor et al (2004a). (a-b) Taux de sélection par session des experts et pendant les 4 essais test des Groupes TP-sep, TP-par, T ; (c-d) Idem pour les groupes DP-sep, DP-par, D.

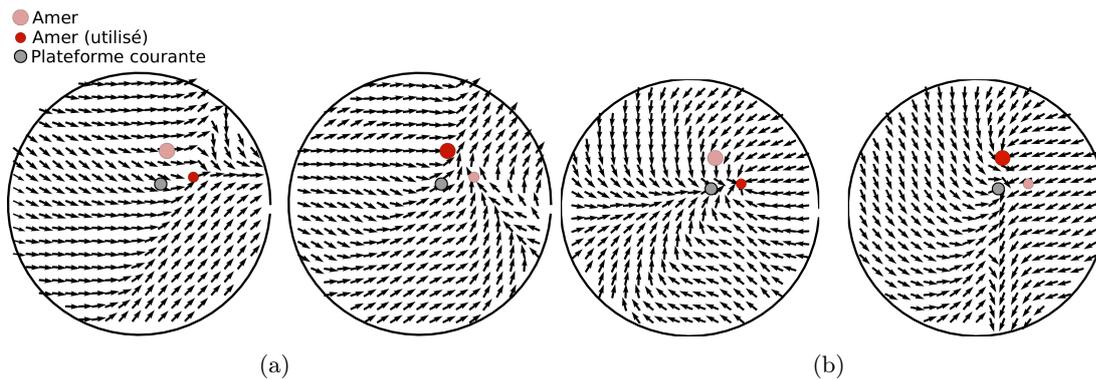


Figure 5.10 – **Simulation de l'expérience de McGregor et al (2004a)** Cartes de navigation des experts (a) Taxon et (b) Direction après l'entraînement.

experts égocentrés apprennent à pointer sur un amer puis cherchent aléatoirement la plateforme, alors que les experts allocentrés pointent vers la région où doit se trouver le but (cf. cartes de navigation des experts de guidage Taxon et Direction, Figure 5.10 a, b). Les deux experts Taxon étant aussi efficaces pour pointer sur un amer, l'un est rapidement sélectionné aux dépens de l'autre. L'amer A est privilégié (il est plus petit, i.e., perçu plus précisément dans le champ visuel de l'animat. Nous reviendrons sur ce point). Les experts Direction allocentrés, eux, coopèrent pour positionner le plus précisément la plateforme par rapport aux amers, d'où leur meilleure performance.

La présence de l'expert Taxon ne favorise pas la reproduction des essais test. La Figure 5.8 c, d montre les temps d'échappement moyens pour rejoindre le but dans les deux essais de test pour tous les groupes. On rappelle que les deux groupes doivent présenter à la fois :

- (i) une absence de différence de temps de latence entre les localisations "opposé" et "correct" en présence d'un seul amer et
- (ii) une différence significative de temps de latence entre les deux localisations en présence des deux amers.

Parmi les groupes constitués de l'expert Taxon, aucun groupe ne répond à ces deux critères à la fois (Figure 5.8 c). Les différences observées entre les temps correspondant aux zones correctes et opposées lors de la présence de deux amers ne sont pas significatives : ils cherchent donc la plateforme au hasard autour de l'amer. En revanche, tous les groupes DP et D y répondent (Figure 5.8 d).

L'occultation de l'un des amers est cause de la difficulté des groupes TP et T pendant les tests. Il a été vu plus haut que, pendant l'entraînement, les groupes TP et T ont (peu) appris la tâche en n'utilisant qu'un seul amer alors que, dans les groupes DP et D, les experts D_A et D_B sont conjointement sélectionnés. Pendant les tests il en est de même. Les groupes D et DP sélectionnent autant D_A et D_B pendant les tests 1-amer et 2-amers (Figure 5.9

d) : alors que l'expert D_A (ou D_B) offre des orientations exploratoires pendant le test 1-amer et localisations "opposées" de la plateforme, il devient utile et coopère avec l'expert D_B (ou D_A) pendant le test 2-amers en localisation "correcte". Ce n'est pas le cas des groupes T et TP, chez qui l'un des experts de guidage joue un rôle mineur (Figure 5.9 b).

5.1.3.3 Discussion partielle

Pendant l'entraînement avec localisation d'un but indicé par deux amers proximaux en l'absence d'indices distaux, lorsque le modèle met en compétition deux experts Taxon égocentrés, leur interaction se traduit par une occultation de l'un des amers. L'apprentissage est peu performant et les essais tests (avec un seul amer et déplacement de la plateforme) peu réussis.

En revanche, la présence d'experts Direction suffit à reproduire la tâche et coopèrent pendant l'entraînement pour positionner la plateforme. Ici l'expert de lieu ne semble pas indispensable pour réussir les tests, on suppose donc que les animaux pourraient choisir seulement une stratégie du type guidage, mais avec un référentiel ancré sur l'amer et non sur la position du corps. Dans le cas du modèle, il s'agit de la sélection alternative de deux experts Direction ayant en entrée chacun un amer. Cette sélection s'avère bien plus efficace que celle d'un seul expert Direction apprenant l'emplacement du but avec un seul amer (cf. Figure 5.6 c), et c'est ce qui explique les différences observées pour retrouver les positions "correctes" de la plateforme.

Les hypothèses de la théorie associative seraient ici étayées : un seul type d'apprentissage par renforcement serait en jeu et les deux experts Direction partageraient la valeur associative des deux amers. La plus faible performance d'un seul expert (cf. résultats avec un seul amer à côté du but) pourrait expliquer le non transfert de l'apprentissage de deux amers à un seul.

McGregor et al (2004a) se sont interrogés sur la nature de la stratégie utilisée par les animaux : soit une capacité d'aller vers les amers et de chercher au hasard autour, soit une capacité "de nager dans une certaine direction des amers". Ils penchent pour la deuxième solution – que nous étayons par nos simulations – se demandant si cette stratégie était toujours utilisée dans le cas d'amer situé à côté du but. Les résultats de la section précédente montrent effectivement que, dans ce cas de figure, cette stratégie est plus efficace qu'un Taxon égocentré.

5.1.4 Bilan de la présence des seuls amers proximaux

Cadre de référence de la stratégie de guidage. Dans cet environnement simulant une piscine de Morris, lorsque le but est visible, les experts de guidage égocentrés et allocentrés ont des performances équivalentes, mais le décalage du but et de l'amer montre que l'expert Direction peut seul apprendre correctement (moins bien que lorsque le but est visible, toutefois) l'emplacement du but. Les groupes DP et D seraient donc privilégiés dans une situation expérimentale où l'amer n'indique pas directement la plateforme.

Pertinence de la stratégie de lieu. Ces simulations ont montré que l'utilisation d'une stratégie de lieu ne peut pas être écartée, dans des conditions où seul un expert de guidage

aurait été jugé compétent, c'est-à-dire en l'absence d'indices distaux. L'expert de lieu positionne le but rapidement mais peu précisément, l'expert de guidage positionne le lieu plus lentement mais précisément. Des travaux récents, avec cependant un nombre d'amers proximaux plus élevé que dans nos simulations, montrent en effet qu'il existent des cellules de lieu réceptives aux amers proximaux, comme il en existe aux amers distaux (Renaudineau et al, 2007). La reproduction de l'expérience de McGregor et al (2004a) a montré toutefois que les groupes avec experts directionnels, en présence de deux amers, pouvaient ne pas utiliser de représentation spatiale.

Reproduction des expériences et prédictions du modèle. Une expérience de localisation d'un but avec deux amers (McGregor et al, 2004a) est ici expliquée avec le concours des seuls experts de guidage. Des rats pourraient donc résoudre cette tâche sans l'aide d'une représentation spatiale. Alors que, de par son apprentissage égocentré, l'expert Taxon occulte immédiatement l'un des amers lors de l'apprentissage, l'expert allocentré Direction partage la valeur associative entre les deux amers. Les deux conditions sont conformes à la théorie associative. Les résultats suggèrent qu'une utilisation d'experts de guidage allocentrés, plutôt qu'égocentrés, pourraient expliquer le comportement des animaux dans l'expérience de McGregor et al (2004a), se fondant sur deux amers proximaux pour positionner la plateforme.

Dans la section suivante, nous n'autorisons la présence que des seuls amers distaux, mais en quantité suffisante pour permettre la construction d'une représentation spatiale correcte.

5.2 Présence des seuls amers distaux

Dans cette section, nous ne prenons en compte que les experts au même référentiel allocentré, c'est-à-dire les experts Direction et Planification, puisque l'expert Taxon égocentré s'est révélé peu pertinent dans le cas d'indices placés à côté du but. Lorsque les entrées sont séparées, l'expert Direction se comporte comme un expert exploratoire. Lorsqu'elles sont partagées, à chaque amer correspond un expert de guidage. L'interaction entre les experts Direction et Planification ne peut donc s'observer pleinement que dans la condition d'entrées partagées.

Nous simulons deux conditions différentes :

- Un apprentissage avec plusieurs amers distaux (en dehors de la piscine), le but invisible étant fixe (inspiré de Morris et al, 1982).
- Un apprentissage avec plusieurs amers distaux, le but invisible étant déplacé (expérience 1 de Steele et Morris, 1999).

Ces expériences fournissent l'occasion de tester l'hypothèse cognitive, c'est-à-dire l'apprentissage d'amers intégrés dans une représentation interne de l'environnement. La deuxième expérience nous permettra d'estimer l'efficacité de l'expert de lieu à s'adapter rapidement à un changement de l'environnement.

Nous intégrons dans les simulations suivantes les deux experts Local et Planification, testés avec le seul expert D. On rappelle que l'expert Local a accès à la même carte de cellules de lieu que

l'expert Planification, mais il ne construit pas de carte topologique. Il apprend simplement les associations lieu-action avec le même apprentissage par renforcement que l'expert Direction et est donc jugé aussi peu flexible que lui.

Des groupes que l'on suppose équivalents aux animaux lésés Hc (D), DLS (L, P) et contrôle (DL, DP) sont simulés avec entrées séparées et partagées.

En présence d'amers distaux - le but étant invisible, non indicé par amers proximaux - l'hypothèse est que l'expert de lieu prenne le contrôle. Il n'est toutefois pas exclu, compte tenu des résultats précédents, que l'expert Direction puisse jouer un rôle, lorsqu'il y aura accès.

5.2.1 But invisible et fixe : expérience inspirée de celle de Morris et al (1982)

5.2.1.1 Protocole et résultats observés

Nous reproduisons les 10 premières sessions de l'expérience de ces auteurs. Dans une piscine de Morris de 1.32 m de diamètre (similaire à celle de la Figure 5.11 a), les auteurs entraînent des rats à rejoindre une plateforme invisible pendant 28 essais. Différents groupes de rats sont testés, dont des rats intacts et des rats hippocampiques. Ils montrent que les courbes d'apprentissage des deux groupes sont significativement différentes : les rats lésés apprennent beaucoup plus difficilement à rejoindre le but (Figure 5.11 b). Ils en déduisent que l'utilisation d'une représentation interne dépendante de l'hippocampe améliore l'apprentissage de la tâche.

5.2.1.2 Protocole et Résultats simulés

Notre environnement est une piscine de Morris de 2 m de diamètre. Il n'y a pas d'amer proximal, mais 8 amers distaux qui sont placés pseudo-aléatoirement autour de la piscine (Figure 5.11 a). Au début de l'expérience, le but est placé à un emplacement qui reste fixe tout le long de l'expérience.

L'expert Planification construit sa carte avec les 8 amers. Lorsque les entrées sont séparées, la condition est équivalente à la simulation d'un seul expert Planification avec des propositions d'exploration dues au bruit reçu par l'expert Direction et l'exploration. Lorsque les entrées sont partagées, 8 experts Direction vont chacun être ancrés sur un amer distal.

Comme dans l'expérience originale, tous les groupes apprennent la tâche. Les groupes DL-sep et DL-par sont les plus performants. En fin d'expérience, tous les groupes obtiennent des performances entre 150 et 30 pas de temps (Figure 5.12 a, b). Les groupes DP-par et P, d'une part, et DP-sep et D, d'autre part, ont un même niveau d'apprentissage. Ils apprennent rapidement et, vers la quatrième session, leurs performances n'augmentent plus.

Les temps d'échappement des groupes avec expert Local sont beaucoup plus faibles que ceux des groupes avec Planification. Cet expert procédural a en effet une capacité d'amélioration de son apprentissage plus importante que l'expert déclaratif, mais il a des performances moins bonnes en début d'expérience. Les groupes DL-sep et DL-par ont les mêmes résultats, meilleurs que ceux des deux groupes L et D. Ils pourraient donc y avoir *coopération* entre ces deux experts, selon la définition de ce type d'interaction.

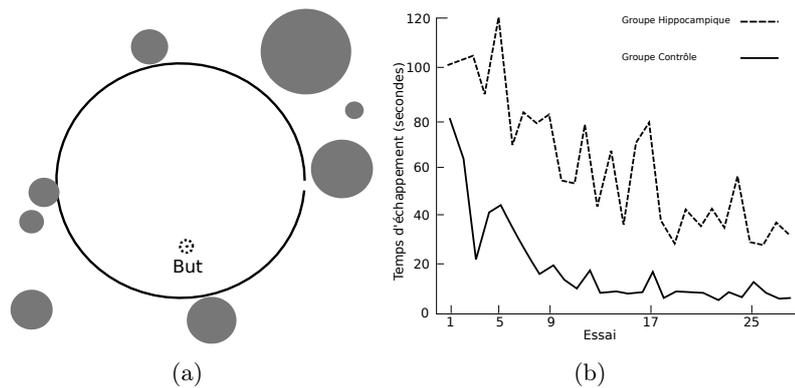


Figure 5.11 – **Expérience inspirée de celle de Morris et al (1982)**. (a) Environnement simulé et placement des amers distaux. (b) Résultats originaux des animaux Contrôle et Hc (Figure 1 de Morris et al, 1982)

Les experts de lieu prennent le contrôle pendant toute l'expérience. Confrontés à un expert Direction, l'expert Local prend le contrôle très majoritairement, comparé à l'expert Planification (Figure 5.12 c, d). Dans les groupes avec expert Local, les experts D et exploration sont sélectionnés en début d'expérience, plus dans le cas des groupes à entrées partagées, où ces experts contribuent à améliorer les performances.

Les performances des groupes DP-sep (où l'expert Direction ne joue qu'un rôle d'exploration) et D sont similaires, pourtant l'expert Planification n'y est pas aussi performant que lorsqu'il est sélectionné uniquement avec de l'exploration. L'expert de lieu prend le contrôle pendant toute l'expérience, mais le taux de sélection d'Exploration et de Direction ne sont pas négligeables (Figure 5.12 d). On peut en déduire que ces deux experts exploratoires perturbent l'expert de lieu et ralentissent ainsi la performance de l'animat. Le groupe DP-par (où l'expert Direction peut aussi s'ancrer sur les amers distaux) a d'aussi bonnes performances que le groupe P. L'expert Planification a un taux de sélection plus important que dans la condition DP-sep (Figure 5.11 d). L'expert Direction n'offrant pas d'orientations aléatoires (car on constate qu'il peut apprendre seul la tâche), il ne perturbe pas l'expert de lieu dans son apprentissage. Sa propre sélection augmente même au long de l'expérience. Il est possible qu'il prenne le contrôle si le nombre d'essais se prolonge, comme supposé par, e.g., Packard et McGaugh (1996).

Dans le groupe D, l'expert Direction peut apprendre la tâche avec des amers spécifiques. Le groupe D, constitué de 8 experts Direction ancrés sur les amers distaux, apprend à positionner la plateforme en s'ancrant *successivement* (et non de façon intégrée, comme l'expert de lieu) sur certains amers.

Les Figures 5.12 e, f illustrent l'évolution des valeurs sélectives des amers (i.e., valeurs que calcule le sous-module de sélection pour chacun des experts Direction) dans les groupes D et DP-par (e, f).

Nous avons mis en annexe les endroits où ces amers ont été choisis pour les positions cardi-

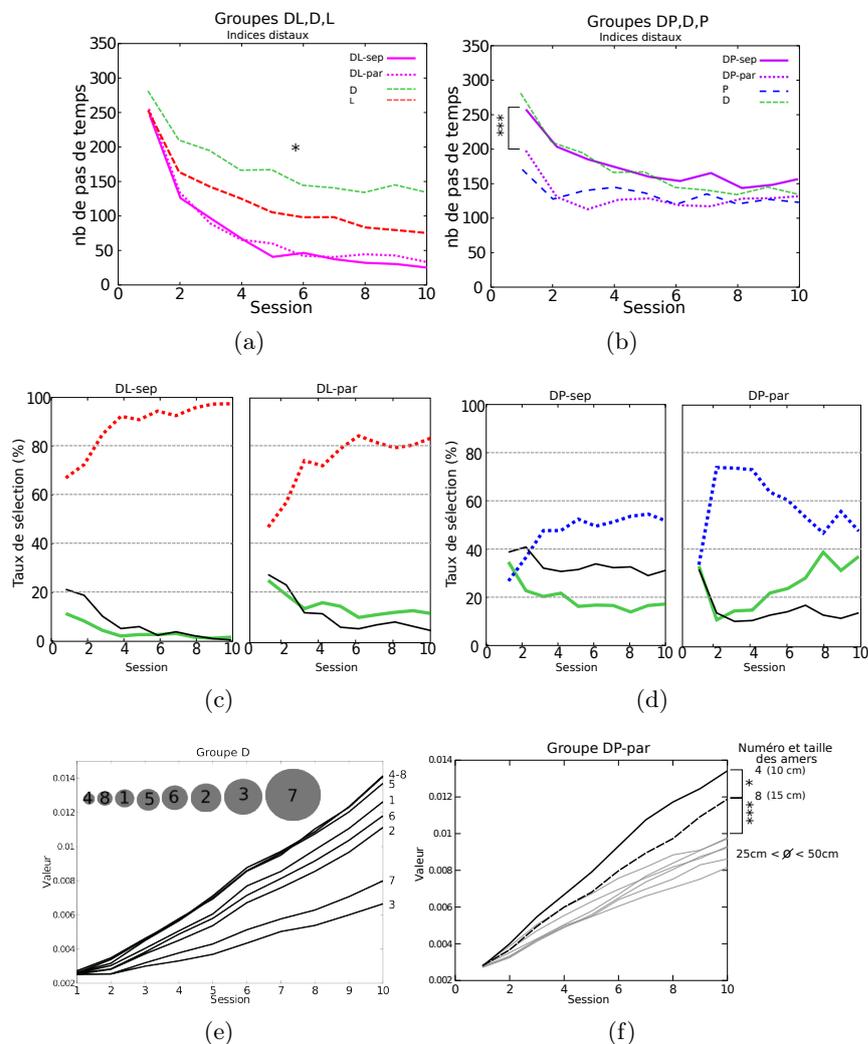


Figure 5.12 – **Expérience inspirée de celle de Morris et al (1982)**. (a) Performances des groupes D, L, DL-sep, DL-par et (b) des groupes D, P, DP-sep, DP-par. (c-d) Taux de sélection des experts par session des groupes précédents. (e) Groupe D : évolution des valeurs de sélection des 8 amers calculées par le module de sélection des experts Direction. (f) Idem pour groupe DP-par

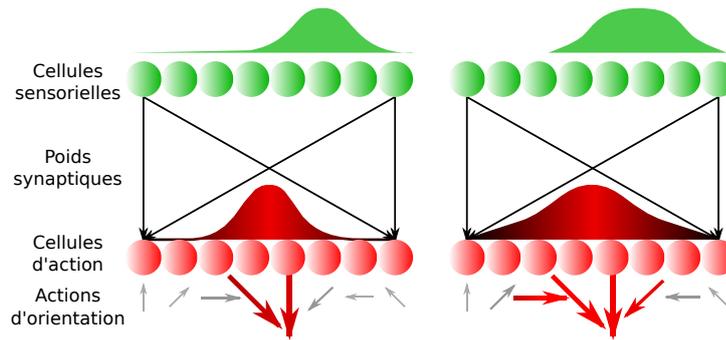


Figure 5.13 – Schéma du renforcement des directions en fonction de la taille d'un amer : plus l'amer est grand, plus nombreuses seront les direction renforcées et moins précise sera la direction finale.

nales de la plateforme (Figure C.1) .

Ces Figures révèlent :

(i) que les amers ne sont pas tous aussi prégnants (cf. l'évolution des poids). Dans nos simulations en effet, les amers les plus petits sont privilégiés (e.g., amers 1, 4 et 8). L'apprentissage de l'expert est en effet d'autant plus précis que l'amer prend peu de place dans le champ visuel de l'animat (remarque déjà indiquée pour l'expert Taxon dans la section précédente). Ceci est en effet dû au fait que la relation entre la taille de l'amer et du renforcement des actions est linéaire (Figure 5.13) : un amer plus petit permettra une navigation plus précise. Ce biais concernant le traitement des informations sensorielle sera repris en Discussion générale.

(ii) que les amers utilisés pour s'orienter vers la direction du but ne sont pas toujours ceux qui sont placés à proximité de la zone où se trouvent les animats (Figure C.1 : par exemple, en haut à gauche, l'amer 6 est utilisé lorsque les animats sont loin ou près de cet amer) ni dans la direction du but. La Figure 5.14 suggère que ces amers pourraient être choisis à la fois pour leur taille et pour leur position par rapport à la trajectoire allant du point de départ de l'animat vers le but.

Certains experts directionnels sont donc successivement sélectionnés pour apprendre la tâche. La valeur associative des amers correspondants augmente pendant l'expérience, mais certains acquièrent plus de saillance que d'autres.

Lorsqu'un expert Planification est présent, il y aurait une sélection plus importante entre experts directionnels. La Figure 5.12 montrent que deux amers (les plus petits) ont une valeur sélective beaucoup plus élevée que les autres : une sélection entre experts directionnels semble plus spécifique dans cette situation. Lorsque les deux experts de guidage et de lieu sont présents et partagent les mêmes indices, les experts de guidage les plus efficaces seraient donc privilégiés, face à la compétence de l'expert Planification.

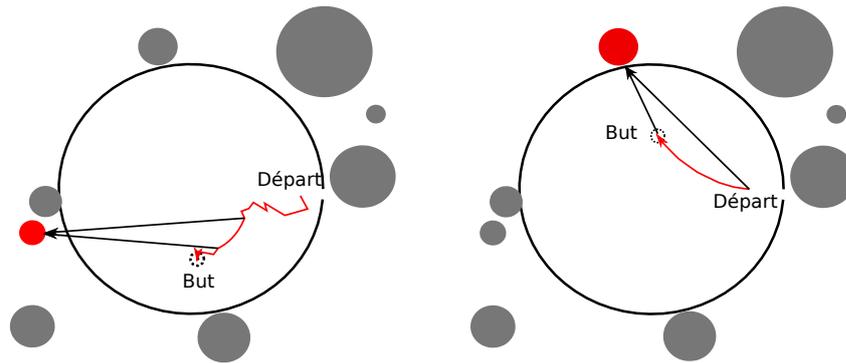


Figure 5.14 – **Expérience inspirée de celle de Morris et al (1982)** . Exemples de trajectoires où l'expert Direction (trajectoire entre les flèches) dirige l'animat vers le but en se basant sur le plus petit amer distal (disque rouge) situé dans la direction du but, par rapport à son point de départ.

5.2.1.3 Discussion partielle

Les hypothèses de départ sont bien confirmées car c'est l'expert de lieu qui est majoritairement choisi pour rejoindre un but invisible indiqué par des amers distaux. Cependant on note qu'un expert de guidage directionnel peut apprendre la tâche, en privilégiant successivement quelques amers particuliers dans l'environnement. Il se révèle moins performant qu'un expert de lieu procédural, mais tout aussi performant que l'expert Planification, avec un entraînement plus important (vers la sixième session, dans l'expérience présente).

Lorsque plusieurs amers sont à disposition de l'expert Direction, certains prennent une valeur associative plus importante que d'autres. Cette situation peut être interprétée en terme d'occultation entre amers, car certains sont plus utilisés que d'autres, mais la sélection conjointe de plusieurs experts directionnels peut aussi assurer un bon apprentissage. Cela avait déjà été évoqué dans l'analyse des résultats de la simulation de l'expérience de McGregor et al (2004b), où deux experts directionnels étaient utilisés conjointement pour positionner la plateforme.

Les performances de l'expert Local, par rapport à l'expert Planification, sont remarquables. En comparant les courbes observées d'animaux Contrôle et hippocampiques et les courbes simulées des groupes DL, DP (équivalents Contrôle) et D (équivalents Hc), on peut même s'interroger sur la nature de la stratégie utilisée par les animaux Contrôle de l'expérience de Morris et al. (Figure 5.11 a). Au lieu de l'utilisation d'une stratégie Planification utilisant une carte topologique – comme supposée par les auteurs –, nous montrons qu'une coopération entre deux stratégies de guidage et de lieu procédurales (Direction et Locale) semble plus conforme au comportement des animaux intact. La stabilité de l'environnement (plateforme et amers distaux fixes) aurait pu en effet inciter les animaux à opter pour ces stratégies, moins coûteuses computationnellement.

L'utilisation non systématique d'une carte cognitive est l'hypothèse qui est au centre de nom-

breux travaux d'Hamilton et al. (e.g., Hamilton et al, 2007, 2008, 2009), suggérant que la stratégie de lieu utilisée par les animaux peut être simplement une stratégie apprenant une direction approximative du but plutôt qu'une localisation absolue de la plateforme. Ils montrent aussi que l'utilisation de cette stratégie pourrait même être préférée à la seconde ou augmenter lorsque les essais se prolongent.

Nous suggérons que la stratégie qu'ils proposent serait plutôt similaire à notre expert Direction, mais qui serait celle utilisée par les rats hippocampiques de Morris et al (1982), et non par les rats Contrôle. Celle qui serait exprimée par ce groupe serait plutôt une *utilisation conjointe de Direction et de Local*. Si cela peut remettre en question l'utilisation d'une carte topologique, cela n'exclut pas, comme le pensent les auteurs, une implication de l'hippocampe dans la réalisation de cette tâche.

5.2.2 But invisible déplacé : expérience inspirée de celle de Steele et Morris (1999)

L'expérience de cet article avait pour objectif d'étudier l'effet de lésions hippocampiques sur l'encodage et le rappel d'un apprentissage spatial dans lequel la plateforme change de localisation à chaque session. L'hypothèse repose sur l'utilisation d'une carte cognitive.

Comme indiqué précédemment, cette condition va permettre de tester la flexibilité de l'expert Planification. Pour estimer cette flexibilité, les performances de cet expert sont comparées avec les performances de l'expert Local. Quatre groupes vont être simulés, DP et DL avec entrées séparées ou partagées. Dans les groupes à entrées séparées, les experts Direction apporteront uniquement de l'exploration aux experts Planification ou Local. Un groupe D constitué de 8 experts directionnels est ajouté pour estimer aussi la flexibilité de la stratégie de guidage.

Compte tenu de son mode d'acquisition, un ré-apprentissage de la tâche est nécessaire aux experts Local et Direction à chaque déplacement du but, leurs cartes de navigation étant à reconstruire à chaque session. L'expert Planification, ayant appris à construire une carte topologique indépendamment du but, pourra plus rapidement résoudre la tâche, une fois qu'il aura pu repositionner le but dans sa carte. Notre hypothèse est donc que l'expert Planification réussira mieux la tâche que l'expert Local. Ce dernier, ayant une carte de cellules de lieu construite avec une intégration d'amers, sera supposé plus efficace que l'expert Direction.

5.2.2.1 Protocole et résultats observés

La piscine a 2 m de diamètre et la plateforme invisible 11 cm de diamètre. Il y a 9 sessions de 4 essais. À chacune des sessions, la plateforme est déplacée aléatoirement dans la piscine à 9 emplacements différents. Les résultats obtenus par Steele et Morris (1999) indiquent que, si l'apprentissage est inexistant entre sessions, les animaux améliorent significativement leur performance à l'intérieur d'une session, i.e., lorsque la plateforme reste à la même place, souvent en un seul essai (Figure 5.15).

Les auteurs en déduisent la capacité des animaux à mettre à jour très rapidement leur repré-

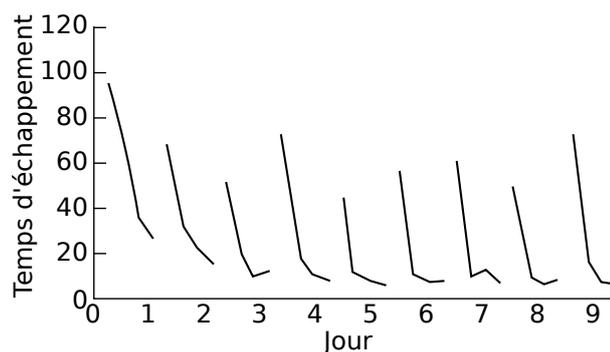


Figure 5.15 – **Expérience de Steele et Morris (1999)** Résultats originaux (voir texte)

sentation spatiale et suggèrent qu'une telle capacité est rendue possible par l'implication de l'hippocampe et la construction d'une carte cognitive. En effet, les rats lésés de l'hippocampe sont uniquement capables de réapprendre la tâche en plusieurs essais, mais non pas en un seul.

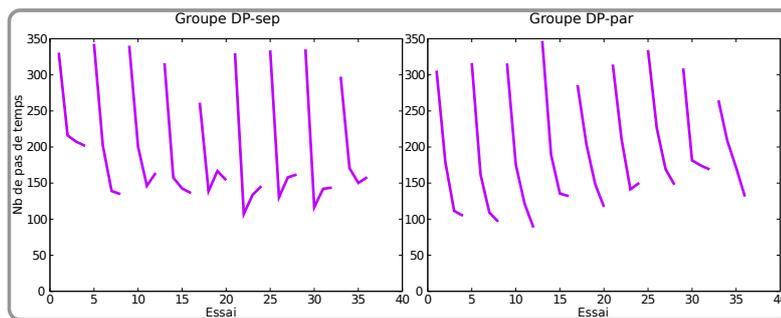
5.2.2.2 Protocole et résultats simulés

L'environnement est le même que celui de l'expérience simulée de Morris et al (1982). Les indices distaux sont au nombre de 8 et placés pseudo-aléatoirement, afin de couvrir toutes les directions autour de l'environnement. L'expérience dure 10 sessions de 4 essais. A chaque session, le but est déplacé à un nouvel emplacement parmi les 8 disponibles. Les temps d'échappement sont relevés dans tous les groupes.

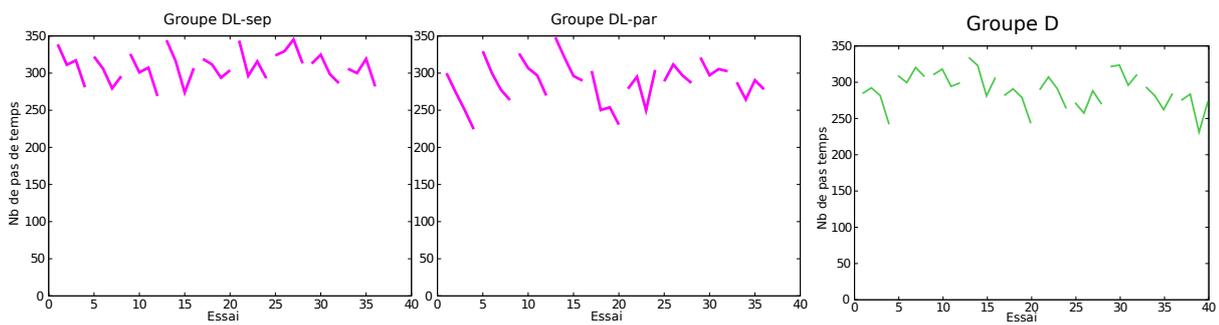
Seuls les groupes avec expert Planification peuvent reproduire la tâche. La Figure 5.16 a montre que, comme les rats contrôles, les animaux du groupe DP (sep et par) réussissent à réapprendre la localisation de la plateforme en un seul essai, en obtenant des niveaux comparables à ceux de la situation précédente, lorsque la plateforme était fixe (entre 80 et 150 pas de temps). Les temps d'échappement aux troisièmes essais, en hausse (non significative), sont dues à l'imprécision de la carte de l'expert Planification. On remarque en effet qu'avec l'aide d'un expert Direction (DP-par), ces effets disparaissent.

En revanche, les groupes DL et D améliorent très peu leurs performances pendant une session : leurs meilleurs temps d'échappement vont au-delà de 250 à 200 pas de temps (Figure 5.16 b, d).

Dans les groupes DP et DL, l'expert de lieu est majoritairement choisi. Les taux de sélection des experts montrent, comme prévu, que les experts de lieu prennent le contrôle rapidement (Figure 5.17 a). Dans le groupe DP-par on ne constate plus, comme dans le cas de la recherche d'une plateforme fixe, une augmentation aussi importante de la sélection de l'expert de guidage en fin d'expérience. La sélection des deux types d'experts n'améliore pas sensiblement les performances. Dans le groupe DL-sep, l'expert Local utilise l'expert Direction et l'exploration en début d'expérience car son apprentissage est plus lent que celui de l'expert



(a)



(b)

(c)

Figure 5.16 – Simulation de l'expérience de Steele et Morris (1999). (a) Courbes d'apprentissage des groupes DP-sep et DP-par ; (b) DL-sep et DL-par ; (c) du groupe D.

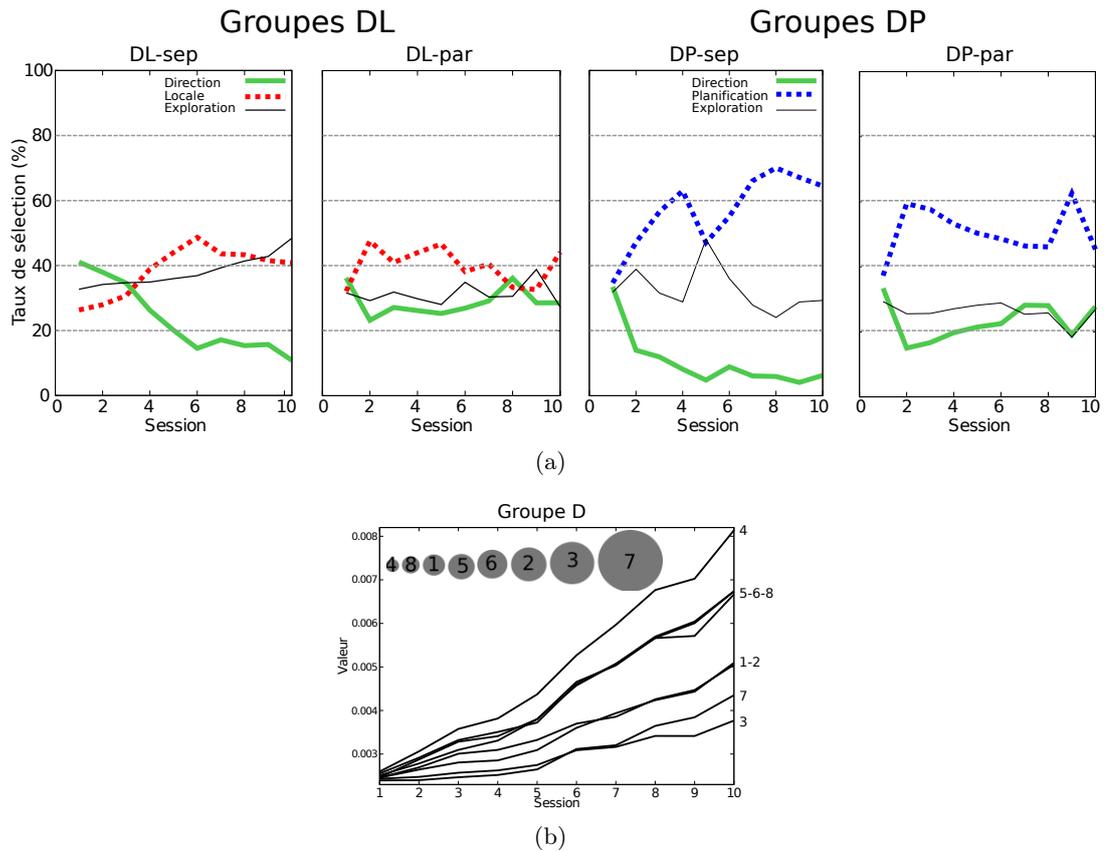


Figure 5.17 – **Simulation de l'expérience de Steele et Morris (1999)** (a) Taux de sélection des experts dans les groupes DL et DP. (b) Evolution des valeurs de sélection des 8 amers calculées par le module de sélection des experts Direction dans le groupe D.

Planification. Dans le groupe DL-par, le taux de sélection de l'expert Direction rejoint celui de l'expert Local, et contribue à une légère amélioration de l'apprentissage (Figure 5.17 b).

Le groupe D a un faible niveau de performance en s'ancrant sur un nombre réduit d'amers. On a vu précédemment les compétences des experts Direction à trouver à l'aide d'amers distaux une plateforme invisible fixe. Ils ont la même flexibilité qu'un expert Local, mais ils réapprennent moins bien le nouvel emplacement de la plateforme (Figure 5.16 c).

La Figure 5.17 b indique qu'il y a une compétition entre experts Direction dans ce groupe et que l'amer le plus privilégié est le plus petit de tous, les trois suivants étant aussi parmi les plus petits. Une capacité d'apprentissage avec quelques amers proximaux avait déjà été notée dans la simulation de l'expérience de McGregor et al (2004a).

5.2.2.3 Discussion partielle

Ces résultats montrent que, lorsque le but est invisible mais change de localisation fréquemment, l'apprentissage de la stratégie de lieu par planification rend mieux compte des données comportementales qu'un apprentissage d'associations lieu-action (expert Local). Nous avons déjà cité le travail de Foster et al (2000) (section 3.2.2.2), qui avait démontré qu'un apprentissage réactif d'une trajectoire ne pouvait expliquer la réussite d'une telle tâche. Seule l'utilisation d'une carte métrique pouvait, selon les auteurs, rendre compte des résultats.

Nous démontrons cependant ici que l'utilisation d'une carte *topologique* peut tout aussi correctement expliquer le comportement des animaux.

Ayant une vitesse d'apprentissage aussi lente que celle de l'expert Local, la stratégie Direction est rendue moins efficace par le déplacement régulier du but, lorsqu'aucun amer distal n'est informatif au début de chaque session. Cette situation augmente la compétition entre les amers. Une stratégie directionnelle ne serait donc sélectionnée plus efficacement (cf. section précédente) que lorsqu'elle peut s'ancrer sur des amers distaux stables et après un entraînement plus long.

5.2.3 Bilan de la présence des seuls amers distaux

Pertinence et nature des stratégies de lieu. La reproduction de l'expérience de Morris et al (1982) montre que la présence d'amers distaux et l'absence d'amers proximaux privilégie, comme prévu, la sélection d'un l'expert de lieu, en cas de plateforme invisible fixe ou déplacée. Cependant l'apprentissage de cet expert de lieu serait de *nature procédurale et non déclarative*, contrairement à l'hypothèse proposée par cette expérience classique de la théorie cognitive.

L'expérience de Steele et Morris (1999)(déplacement de la plateforme à chaque session) montre les limites de la stratégie Locale comme implémentation efficace de la stratégie de lieu. La flexibilité requise par cette tâche ne peut en effet être reproduite par un simple *Q-learning*, une observation déjà indiquée par Foster et al (2000).

Nous reviendrons dans la Discussion générale sur la co-existence possible des stratégies Planification et Locale. Une telle distinction pourrait être mieux prise en compte dans l'implication qu'elle pourrait avoir sur la nature de l'apprentissage spatial.

Pertinence de la stratégie de guidage. Il s'avère qu'une telle stratégie peut apprendre, après plusieurs essais, la position d'un but invisible fixe par des amers distaux stables, ce que Morris et al (1982) avaient déjà démontré avec des animaux lésés. Hamilton et al (2008) ont démontré cependant que cette stratégie de direction est aussi adoptée par les rats lorsque la plateforme est déplacée plusieurs fois au cours de l'expérience, résultat que nous n'avons pas obtenu dans ces simulations. Si la stratégie suggérée par Hamilton est plus proche d'une stratégie conjointe Direction/Locale (DL-par), alors en effet un meilleur mais toujours faible apprentissage peut-être observé.

Entrées des experts : séparées ou partagées ? Il semblerait que la condition dans laquelle les experts ont accès à tous les amers soit plus conforme aux observations comportementales lorsque le but est fixe, avec une augmentation de la sélection de la stratégie de direction suggérée

par certains auteurs lorsque l'expérience se prolonge (Hamilton et al, 2008). Lorsque le but est déplacé, il est plus difficile de préciser si l'expert de guidage doit être sélectionné aussi, les résultats des deux conditions n'étant pas suffisamment différents.

Reproduction des expériences et prédictions du modèle. Les deux expériences but fixe (Morris et al, 1982) et but déplacé (Steele et Morris, 1999) en présence des seuls amers distaux, expériences classiques de la théorie cognitive, ont été bien reproduites. Le modèle prédit ce que d'autres expérimentations ont déjà mis en évidence, soit la possibilité pour des rats hippocampiques d'apprendre à positionner un but fixe, et de ne pas pouvoir apprendre à repositionner la plateforme en un seul essai lorsqu'elle est déplacée. Toutefois nos résultats suggèrent, pour des rats intacts, une utilisation distincte de stratégies de lieu dans les deux conditions, une procédurale pour la première, une déclarative pour la deuxième.

5.3 Conclusion du chapitre

Au-delà de la validation du modèle sur un certain nombre d'expériences simples (rejoindre un but visible ou indicé, non visible mais indiqué par des amers distaux...), cette série d'expériences, dont les résultats sont résumés dans le tableau 5.1 avec la présence d'un seul type d'amers, proximaux ou distaux, a permis de dégager les points suivants :

En présence d'amers proximaux et en l'absence d'amers distaux.

- Une stratégie de guidage est plus performante qu'une stratégie de lieu pour rejoindre un but visible ou invisible, indicé par des amers proximaux.
- Dans les environnements où le but est invisible mais indicé par des amers proximaux situés à côté de ce but, une stratégie de guidage égocentrée (Taxon) est significativement moins performante qu'une stratégie allocentrée (Direction)
- La localisation d'un but en présence de deux amers à côté du but est mieux apprise par un expert de guidage allocentré. Dans la situation simulée, un expert égocentré a tendance à occulter l'un des amers, et est moins performant.
- On ne peut écarter la participation d'un expert de lieu, qui peut construire une carte - d'efficacité limitée - dans un environnement pauvre.

En présence d'amers distaux et en l'absence d'amers proximaux.

- Une stratégie de lieu procédurale est plus performante qu'une stratégie de lieu déclarative pour rejoindre un but invisible et fixe, indicé par les seuls amers distaux, après un entraînement plus intensif.
- Une stratégie de type guidage est moins performante que les stratégies de lieu pour rejoindre un but invisible et fixe, indicé par les seuls amers distaux, avec un entraînement plus intensif.
- Cette stratégie de guidage a tendance à s'ancrer non pas sur une intégration d'amers, comme une stratégie de lieu, mais successivement sur certains amers spécifiques.

- En cas de déplacements fréquents du but, seule une stratégie Planification peut apprendre à repositionner un but invisible en peu d'essais. Une représentation spatiale de type topologique et non métrique suffit à réussir cette tâche.

Effets associatifs et cognitifs. Les résultats acquis par le modèle suggèrent que des stratégies apprises par un algorithme de type procédural peuvent rendre compte – tout aussi bien qu'une stratégie reposant sur l'utilisation d'une représentation topologique de l'environnement – de comportements permettant aux animaux de rejoindre une plateforme visible ou indiquée fixe. En revanche, elles ne sont pas assez flexibles pour s'adapter à des environnements changeant fréquemment. Ceci conforterait les hypothèses cognitives.

Par ailleurs, des effets d'occultation entre amers caractéristiques d'un apprentissage associatif simultané ont pu être obtenus par le modèle. En présence d'amers (proximaux, distaux) supposés spécifiques de différents systèmes de navigation, de stratégies (resp. de guidage ou de lieu) peut occulter certains de ces amers (resp. distaux ou proximaux) dès le début de l'expérience. De même, lorsque plusieurs experts effectuent le même apprentissage procédural dans un environnement où de nombreux amers sont présents, une occultation de certains d'entre eux peut résulter de leur interaction. Ceci conforterait les hypothèses associatives.

Le modèle peut donc rendre compte d'effets caractéristiques de l'une ou de l'autre des théories antagonistes. Il a aussi montré – dans une expérience consistant à rechercher une plateforme indiquée par deux amers proximaux, sans amers distaux – que les deux types d'experts de guidage et de lieu ou un seul type, un expert de guidage, pouvaient aussi bien rendre compte du comportement des animaux.

Il est donc possible pour le modèle de suggérer des hypothèses concernant les situations expérimentales qui pourraient étayer l'une des deux théories et non l'autre, et celles qui pourraient être expliquées à la fois par l'une et par l'autre – donnant ainsi une préférence au "principe de parcimonie" (rasoir d'Ockam) consistant à opter pour l'absence d'utilisation d'une carte cognitive.

Enfin, nous notons que toutes les expériences simulées dans ce Chapitre peuvent être aussi bien effectuées par des groupes avec entrées séparées ou partagées.

Dans le chapitre suivant, avec l'appui de ces conclusions, quelques situations environnementales faisant intervenir à la fois des amers proximaux et distaux seront analysées et interprétées en termes d'interactions d'experts de guidage et/ou de lieu.

Tableau 5.1 – Résumé des résultats simulés des groupes Contrôles de ce chapitre, selon les configurations d'amers et si le but est déplacé ou non pendant l'entraînement (Ent.). Les expériences sont indiquées par le nom du premier auteur de l'article correspondant. Les groupes "Conforme" sont ceux qui reproduisent au plus près le comportement animal (ou, dans le cas d'EXP1, EXP2, ceux qui sont les plus performants); les groupes "Possible" sont ceux qui présentent quelques différences avec le comportement observé; les "Non conforme" ne reproduisent pas les résultats. Les groupes "Non simulé" n'ont pas participé à ces expériences. Enfin les sous-catégories "Isol.", "Sep." et "Par." signifient si les simulations ont été réalisées avec, respectivement, un experts seul, des experts avec entrées séparées ou partagées.

Exp.	Amers	But déplacé	Conforme			Non conforme		
			Isol.	Sep.	Par.	Isol.	Sep.	Par.
Exp1	Balise	non	T D	DP				
Exp2	1 prox	non			TP,DP	T	TP	
McGregor	2 prox	non	D		DP	T	TP	TP
Morris	8 dist	non		DL	DL	P D	DP	DP
Steele	8 dist	oui		DP	DP	D	DL	DL

Exp.	Amers	But déplacé	Possible			Non simulé		
			Isol.	Sep.	Par.	Isol.	Sep.	Par.
Exp1	Balise	non		TP	TP,DP	L P	TL,DL	TL,DL
Exp2	1 prox	non	D	DP	TP	L P	TL,DL	TL,DL
McGregor	2 prox	non		DP		P L	TL,DL	TL,DL
Morris	8 dist	non	L			T	TL,TP	TL,TP
Steele	8 dist	oui				T L P	TL,TP	TL,TP

Chapitre 6

Interactions entre amers proximaux ou/et distaux

Ce chapitre décrit des simulations d'expériences dans lesquelles des amers proximaux et distaux peuvent participer simultanément à l'apprentissage spatial, en étant supposés être gérés par un même système. Selon les auteurs, ces apprentissages suivent une règle associative (Rodrigo et al, 2006; Roberts et Pearce, 1999) ou la contredisent (McGregor et al, 2004a). Nous cherchons à montrer dans quelle mesure les mécanismes du modèle multi-expert peuvent permettre de rendre compte de ces différents effets de l'apprentissage spatial.

Les sections suivantes décrivent :

- un effet de gradient associatif du à la distance entre un amer proximal et un amer distal dans l'expérience 1 de Rodrigo et al (2006)
- un effet de blocage d'amers distaux par un amer proximal dans l'expérience 3 de Roberts et Pearce (1999)
- une absence d'occultation de deux amers proximaux par des amers distaux dans l'expérience 2 de McGregor et al (2004a)

Nous ne simulerons pas toujours tous les experts disponibles, mais les choisirons selon les conditions expérimentales.

6.1 Effet de gradient associatif du à la distance entre un amer proximal et un amer distal (Rodrigo et al, 2006)

L'expérience de Rodrigo et al (2006) tend à démontrer l'effet d'un processus associatif dans la relation entre un amer distal F et un amer proximal B en manipulant la distance entre ces deux indices. L'apprentissage du but ayant été appris avec une certaine distance entre les deux amers, les auteurs s'attendent à une perte de confiance dans ces indices lorsque cette distance augmente, avec néanmoins une préférence pour l'amer proximal.

Nous supposons que, si ces résultats se retrouvent dans les simulations, cette perte de confiance pourrait être due à une compétition entre les deux experts de guidage, ou un ex-

pert de guidage et de lieu, entraînant l'animat vers chacun des deux amers. L'expert de lieu ne pouvant construire une carte fiable avec peu d'amers, la préférence pour l'amer proximal pourrait être due à une performance supérieure de l'expert de guidage. Il se pourrait que seul cet expert puisse réaliser la tâche, l'hypothèse étant que, puisque dans nos simulations il est simulé en deux experts séparés s'ancrant chacun sur un amer (proximal ou distal), ils pourraient entrer aussi en compétition.

6.1.1 Protocole et résultats observés

Nous reproduisons l'expérience 1 de Rodrigo et al (2006). Les rats doivent rejoindre dans une piscine (1,6 m de diamètre) une plateforme cachée (10 cm de diamètre), indiquée par un amer proximal B (un cône) situé directement au-dessus d'elle. Un amer distal F (une sphère), se trouvant hors de la piscine, est également présent, situé de sorte que les rats rejoignant la plateforme nagent dans la direction de cet amer. L'amer proximal est éloigné de l'amer distal de 38 cm (Figure 6.1 a). Des rideaux noirs occultent la pièce. Après un entraînement de 4 sessions de 8 essais commencent des sessions de tests.

Ces sessions de test (sessions 6, 8, 10 et 12) sont composées de 8 essais identiques à ceux de l'entraînement, plus un 9^{ème} où la plateforme est enlevée et où l'amer proximal est éloigné de l'amer distal en suivant une rotation de 45 degrés par session (45, 90, 135 et 180 degrés dans le sens horaire ou anti-horaire selon l'individu), par rapport au centre de la piscine (Figure 6.1 a). Ce test vérifie sur quel amer s'ancre l'animal pour s'orienter.

Après chacune des sessions test s'intercale une session d'extinction de l'apprentissage de l'amer distal (sessions 5, 7, 9 et 11), composée de 8 essais identiques à ceux de l'entraînement et un autre essai (qui peut s'intercaler entre les 8 autres) pendant lequel la plateforme et l'amer B sont enlevés. Selon les auteurs, ces sessions ont pour but de réduire la possibilité d'occultation de l'amer distal par l'amer proximal pour mettre en valeur l'effet de gradient de distance. Les auteurs mesurent les temps d'échappement et estiment l'effet du gradient en notant le temps passé par les rats dans l'octant de la piscine (la portion du disque représentant 1/8^{ème} de sa surface) correspondant à l'amer proximal et celui correspondant à l'amer distal.

Les rats améliorent leur performance durant l'entraînement, mais pas lors des sessions tests et extinctions (Figure 6.1 b). Les résultats montrent un gradient d'occupation de zone de l'amer proximal B : plus il est éloigné de sa position d'origine, moins les animaux occupent son octant (les différences sont non significatives pour les deux derniers déplacements, Figure 6.1 c, gauche). Cependant il n'y a pas de gradient correspondant pour la zone de l'amer distal F : dès que l'amer proximal B est éloigné, la zone de l'amer distal F n'est plus occupée, suggérant ainsi une brusque perte de confiance en cet amer (Figure 6.1 c, droit).

Les auteurs en déduisent :

- (i) qu'il y a effectivement un gradient de valeur associative proportionnel à la distance entre l'amer proximal et l'amer distal : plus la distance augmente, moins l'amer proximal serait saillant.
- (ii) qu'il y a quand même une occultation de l'amer F par l'amer B, donc que l'amer proche

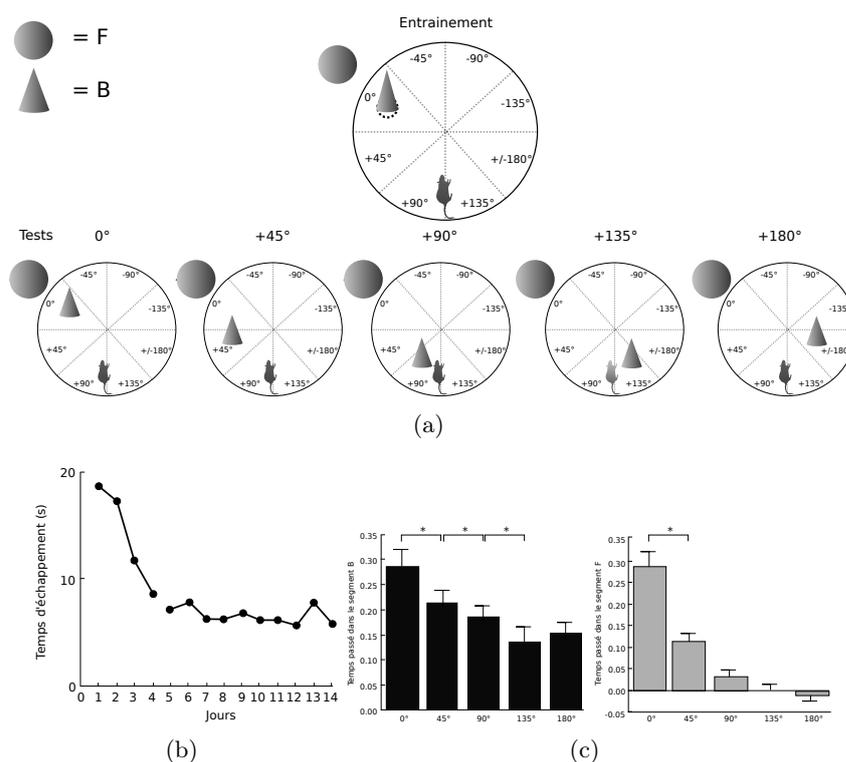


Figure 6.1 – **Expérience 1 de Rodrigo et al (2006)** (a) Protocole original (voir texte) (b) Résultats : apprentissage durant les 14 sessions. (c) Résultats : gradient du temps d'occupation de l'octant B pour chaque déplacement de l'amer et absence de gradient du temps d'occupation de l'octant F. Les temps d'occupation liés au hasard sont soustraits des temps effectifs, ce qui explique qu'il soit possible d'obtenir des valeurs négatives au final.

du but est plus prégnant. Deux autres expériences confirment ces conclusions, les auteurs ayant écarté d'autres facteurs, dont l'effet de la locomotion, qui aurait été plus longue vers les amers les plus éloignés.

6.1.2 Protocole et résultats simulés

Les dimensions de la piscine et de la plateforme sont équivalentes à celles de l'expérience originale. Pendant l'entraînement, la plateforme et l'amer proximal situé sur la plateforme sont placés à 38 cm de l'amer distal. Nous suivons le même protocole que dans l'expérience pour les sessions de test et d'extinction. Les points de départ des animats sont situés aux quatre points cardinaux, tirés au sort pour qu'il y en ait le même nombre.

L'amer proximal étant placé sur le but, nous intégrerons aussi des groupes avec expert Taxon. Les groupes suivants seront simulés :

- un groupe aux entrées séparées TP-sep (où l'expert T est ancré sur B et l'expert P construit

- sa carte avec F)
- un groupe aux entrées partagées TP-par (où l'expert T est ancré sur B et sur F et l'expert P construit sa carte avec B et F)
- un groupe T-par, constitué des deux experts Taxon (ancrés chacun sur un amer). Ce groupe renseignera la nécessité ou non d'une construction de carte cognitive.
- trois autres groupes où un expert Direction remplace l'expert Taxon ci-dessus.

L'intervention d'un expert Local a été écartée, suite au peu d'essais disponibles lors des déplacements de plateforme.

Comme les auteurs, nous mesurons les temps d'échappement et le temps passé dans les octants près de l'amer proximal B (octant B) et près de l'amer distal F (octant F) dans les essais tests (sessions 6, 8, 10, 12) et près de l'amer F dans les tests d'extinctions (sessions 5, 7, 9, 11). Pour analyser les sélections d'experts pendant les tests, nous calculons la moyenne de leurs taux dans les 5 essais de test (amer proximal déplacé et plateforme absente) puis, pour détailler encore plus, (i) avant que les animats atteignent l'octant B pour la première fois et après, pendant qu'il se trouvent, (ii) dans l'octant B et (iii) dans les autres octants.

Ces taux de sélection reflètent aussi le temps passé dans ces différentes phases, respectivement : (i) le temps que les animats ont mis *avant* d'atteindre pour la première fois l'octant de l'amer déplacé, puis le temps passé, (ii) dans l'octant B et (iii) dans tous les autres octants *après* jusqu'à la fin de l'essai.

Nous relevons également quelques trajectoires individuelles pour illustrer nos interprétations des résultats.

6.1.2.1 Le groupe DP-sep est le plus conforme au comportement des animaux

Les courbes d'apprentissage de tous les groupes sont conformes aux résultats observés. Pendant l'entraînement, les performances s'améliorent rapidement et ne diminuent plus lors des sessions test et extinction (Figure 6.2 a, b). Il n'y a pas de différences significative de performance pendant l'entraînement. On note cependant que les groupes DP présentent une amélioration de performance pendant l'entraînement plus conforme à la courbe d'apprentissage originale.

Pendant les 5 essais de test, les effets de gradient des 4 premiers octants B (de 0° à 135°) et d'occultation de F sont retrouvés dans le groupe DP-sep. Les effets de gradient dans les taux d'occupation des octants B et F pendant les essais de test (Figure 6.2 c, e, f) ne se retrouvent pas dans le groupe dans lequel deux experts Taxon s'ancrent chacun sur un amer (T-par, 6.2 c, droit).

Une différence significative entre des taux d'occupation des quatre premiers octants B (similaire à ceux des animaux, de moins en moins importants) ne se retrouve en revanche que dans les groupes DP-sep et D-par (Figure 6.2 e, f droit, comparé à c).

Pendant les 5 essais d'extinction, seul le groupe DP-sep a un taux d'occupation dans l'octant F différent du hasard. On rappelle que les auteurs avaient effectué des

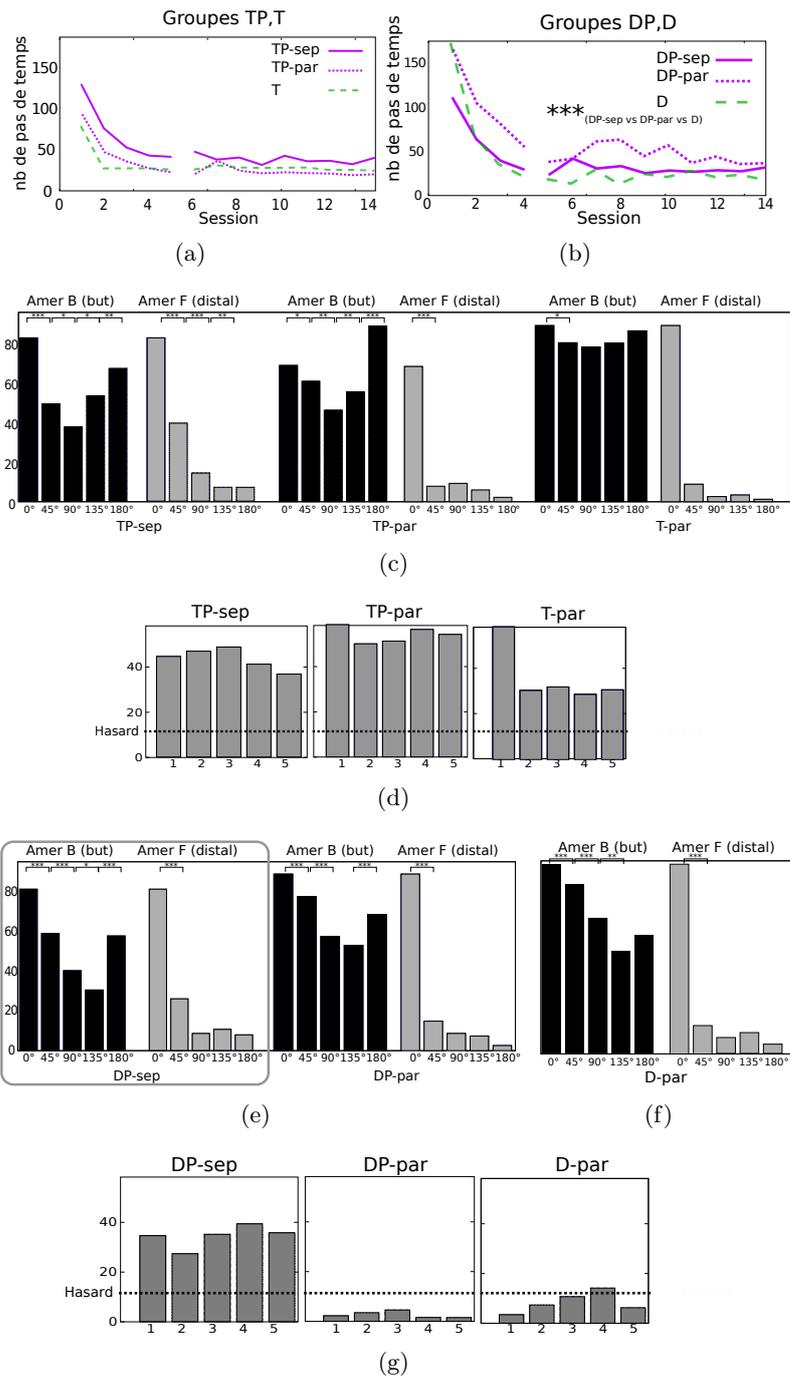


Figure 6.2 – **Simulation de l'expérience de Rodrigo et al (2006)** : (a-b) Performances des groupes durant les 14 sessions. (c,e,f) Essais test. Taux d'occupation dans les octants B et F lors des 5 essais tests (un seul essai par orientation) pour (c) les groupes TP-sep, TP-par, T-par ; (e) le groupe DP-sep ; (f) les groupes DP-par et D-par. (d, g) : Essais d'extinction (atténuation de l'occultation de l'amer distal) : temps passé dans l'octant F lors des 5 essais d'extinction, sans amer B ni plateforme pour (d) les groupes avec expert Taxon et (g) les groupes avec expert Direction.

essais d'extinction de l'amer B afin que l'occultation de F par B ne soit pas trop importante. Ils avaient en effet constaté que les animaux présentaient un taux d'occupation de l'octant F différent du hasard lors de ces essais, lorsque seul l'amer distal est disponible. Cela signifiait que les animaux avaient pu apprendre une stratégie de lieu pendant les 8 autres essais de ces sessions (en présence de F, B et de la plateforme) et justement réduire, comme prévu, l'occultation de F.

Les taux d'occupation de cet octant chez les animats pendant ces essais démontrent que seul le groupe DP-sep est conforme à ce critère (Figure 6.2 g ; $p < 0.001$).

Considérant les trois critères précédemment cités :

(i) des courbes d'apprentissage (différences plus importantes de performance entre l'entraînement et les sessions tests),

(ii) des gradients descendants significativement différents entre les 4 premiers déplacements pour le taux d'occupation de l'octant B, une chute de confiance pour l'amer F dans les essais test

(iii) un taux d'occupation de l'octant F significativement différent du hasard dans les essais d'extinction,

le groupe qui paraît le mieux réunir ces conditions est le groupe DP-sep. Les groupes TP-sep et TP-par les remplissent dans une moindre mesure, au vu de leurs différences de gradients d'occupation moins conformes.

On note donc *qu'aucune des deux conditions d'experts Taxon ou d'experts Direction seuls ne correspondent aux résultats observés.*

Nous allons d'abord analyser les résultats du groupe DP-sep, puis ceux des autres groupes cités (les Figures correspondants aux groupes avec expert Taxon ont été mises en annexe D).

Nous allons analyser plus précisément comment les experts contribuent au gradient d'occupation des octants B (à amers déplacés) .

6.1.2.2 Analyse du groupe DP-sep

Pendant l'entraînement. Après une sélection conjointe avec l'expert Planification, l'expert Direction B prend le contrôle (Figure 6.3 a). Notons que dans les groupes à entrées partagées (DP-par, D), l'expert Direction F est aussi dévalué (Figure 6.3 b, c). L'effet d'occultation de l'amer distal F indiqué par les auteurs peut donc se voir dès l'entraînement.

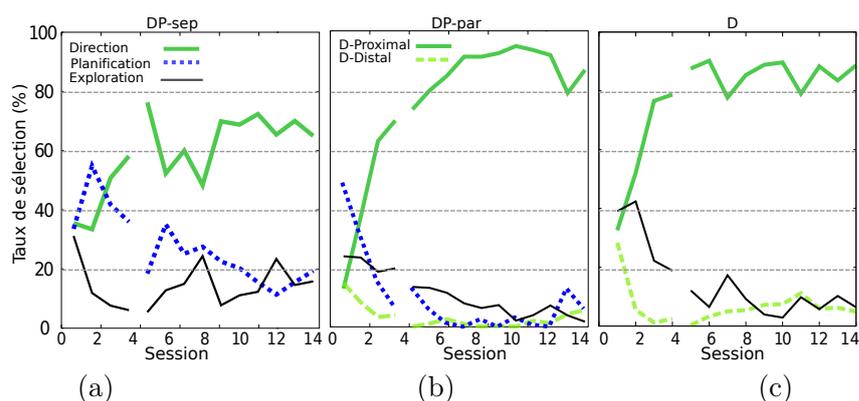


Figure 6.3 – **Simulation de l'expérience de Rodrigo et al (2006)** : (a-c) Taux de sélection des experts des groupes DP-sep, DP-par et D durant les sessions d'entraînement (1 à 4), de test et d'extinction (5 à 14).

Dans les essais tests. Ces résultats concernent la moyenne des 9 essais de chaque session et non uniquement les cinq 9^{èmes} essais pendant lesquels l'amer proximal est déplacé. Tous essais confondus, l'ancrage sur l'amer proximal se confirme (Figure 6.3 a). Le taux de sélection de l'expert de lieu diminue : l'amer F devient moins saillant.

Les taux cumulés des sélections d'experts dans les essais test sont similaires pour tous les octants. La Figure 6.4 a montre effectivement que (contrairement aux groupes DP-par ou D : Figure 6.4b, c), les animaux sélectionnent autant les deux experts et l'exploration pour l'occupation de tous les octants. Dans le groupe DP-sep, le gradient d'occupation ne se traduit donc pas par une graduation dans le choix des experts ou de l'exploration dans les taux cumulés de sélection d'experts pour les différents octants.

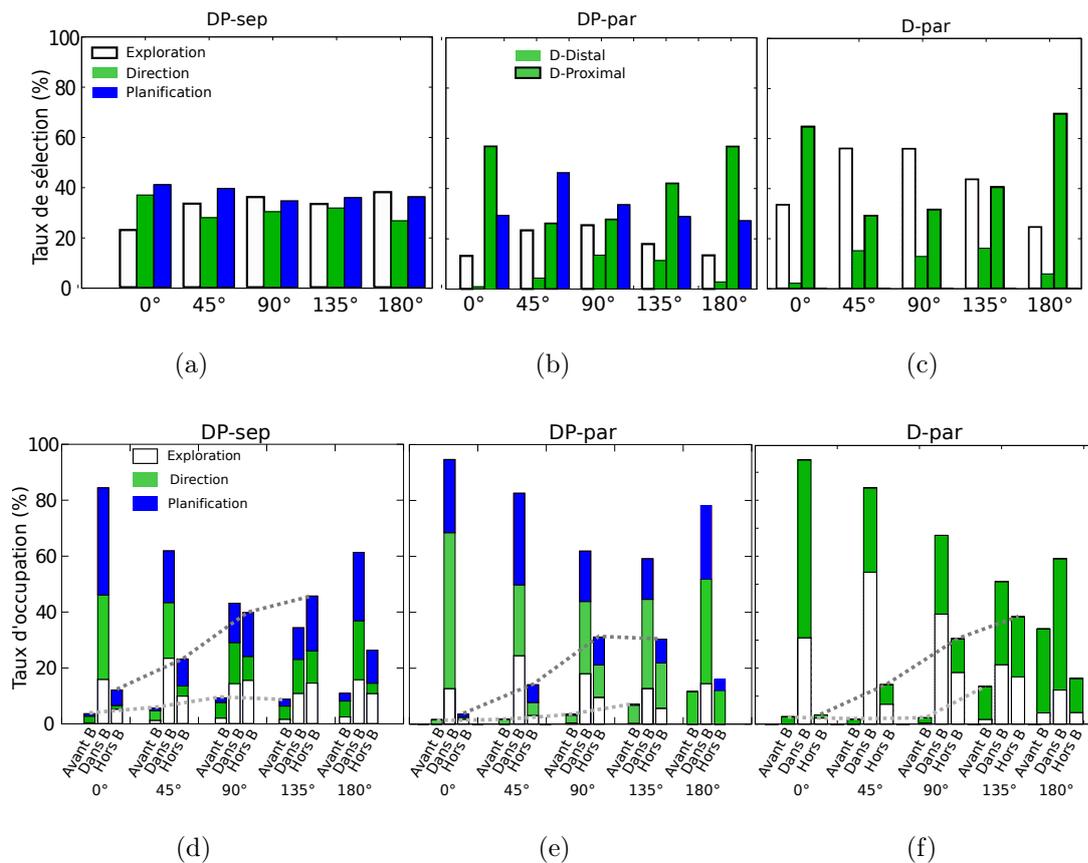


Figure 6.4 – **Simulation de l’expérience de Rodrigo et al (2006)** : Groupes DP-sep, DP-par et D durant les sessions d’entraînement (1 à 4), de test et d’extinction (5 à 14). (a-c) Comparaison des taux de sélection des experts pendant les 5 essais test correspondant à chacune des 5 orientations. (d-f) Taux de sélection des stratégies pondérés par le temps passé pendant les 5 essais test de déplacement de l’amer B (sans plateforme) *avant* d’atteindre l’octant B pour la première fois et après, *dans* cet octant et *hors* de cet octant.

Sans effet de gradient, les animaux rejoignent très rapidement les octants B. Détaillant encore plus le choix des animaux pour différentes Phases des essais tests, la Figure 6.4 d (voir Figure 6.4 e, f pour les autres groupes) sépare les sélections *avant* d’atteindre l’octant B pour la première fois (premières barres) et *après*, dans l’octant B (deuxièmes barres) et hors de cet octant, octant F compris (troisièmes barres). Les premières barres de la Figure 6.4 d montrent que les animaux parviennent tout aussi rapidement dans les différents octants de l’amer B. Le taux de sélection de l’expert Direction est ici majoritaire : l’amer proximal B est très prégnant et l’expert Direction B est efficace pour mener l’animal vers le bon octant.

L’effet de gradient serait dû à la nouvelle recherche de la plateforme. Lorsque les animaux arrivent à l’endroit où peut se trouver la plateforme, ils ne reçoivent pas de récompense.

Ils déclenchent alors une recherche de la plateforme par leurs différents experts. Les deuxièmes et troisièmes barres de la Figure 6.4 d montrent que les trois experts – Planification, Direction, Exploration – sont recrutés pour chercher la plateforme avec une sélection conjointe, descendante pour les occupations de l’octant B, ascendante pour l’occupation des autres octants, pour les quatre premiers déplacements de l’amer B. Ceci est conforté par la Figure 6.5, montrant des trajectoires typiques dans lesquelles l’animat se dirige tout d’abord rapidement vers la plateforme en sélectionnant l’expert Direction B (Figure 6.5, trajectoires a et b). Une fois l’amer B rejoint, l’expert Planification va entraîner l’animat hors de l’octant B (Figure 6.5, trajectoire a) ou parfois jouer le rôle d’exploration (Figure 6.5, trajectoire b). D’après sa carte en effet, l’expert Planification n’a mémorisé l’emplacement de la carte qu’avec l’amer F, puisqu’en B il n’a pas reçu de récompense (Figure 6.5, cartes a ou b).

En résumé, l’expert Direction B serait celui qui guiderait majoritairement les animats vers la plateforme (vers l’amer B), puis une compétition avec l’expert Planification, accompagné d’exploration, se mettrait en place. La compétition augmenterait au fur et à mesure des éloignements de l’amer B (cf. taux de sélection de Planification augmentant dans les autres octants de 4 à 20%) – entraînant de plus en plus les animats hors de l’octant B pour les quatre premiers déplacements d’amer.

Pendant ces essais, l’occultation de l’amer F est dû à une plus grande efficacité de l’expert Direction B. L’octant F est faiblement occupé dès le premier essai de déplacement. L’expert Planification entraîne les animats hors de l’octant B, mais pas jusqu’à l’octant F. D’une part, cet expert est utilisé pour explorer l’octant B, comme signalé précédemment. Lorsqu’il entraîne les animats hors de cet octant, la trajectoire est corrigée par l’expert Direction B, qui est plus efficace (Figure 6.3 a), ce qui est cause de la faible occupation de l’octant F.

L’expert de lieu contribue à l’occupation de l’octant F dans les sessions d’extinction. Pendant l’entraînement, l’expert Planification du groupe DP-sep a été plus efficace que celui du groupe DP-par, car il a plus contribué à l’amélioration de l’apprentissage (Figure 6.3 a, comparé à b). Le module de sélection de ce groupe a donc tendance à lui accorder plus de confiance à l’issue de l’entraînement. Les animats ont suffisamment atténué leur effet de blocage de l’amer distal (constaté dès le début de l’entraînement) pour augmenter leur taux d’occupation de l’octant de la plateforme dans les sessions d’extinction.

6.1.2.3 Sélection des experts dans les groupes sans expert Planification

Nous avons mentionné que les groupes avec les seuls experts de guidage ne pouvaient pas représenter le comportement des animaux.

Le groupe D-par présente l’effet de gradient, pas d’extinction. Il a été signalé que les animats comprenant des experts de guidage seulement ne reproduisaient pas totalement le comportement des animaux. Il est vrai que le groupe avec deux experts Direction reproduit le gradient de taux d’occupation observé (Figure 6.2 f droit). Cet effet est dû à une sélection de plus en plus importante de l’expert Direction F hors de l’octant B (Figure 6.4 c, f). Cependant

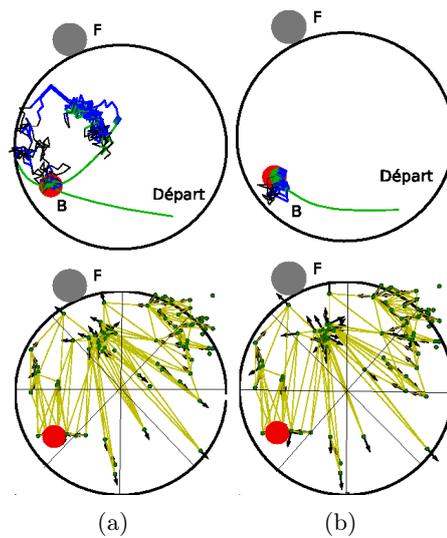


Figure 6.5 – **Simulation de l'expérience de Rodrigo et al (2006)**. Haut : Exemples de sélection d'experts pour un animat DP-sep pendant une trajectoire lorsque B (en rouge) est orienté à 90° . Notez la première Phase de la trajectoire (a,b), menant rapidement au but avec l'apport du seul expert Direction B, puis (a) la sélection de l'expert Planification entraînant l'animat vers F ou (b) explorant autour de la position de B.

les essais d'extinction de l'amer F n'ont pas eu d'effet sur eux (6.2 g droit). Les taux de sélection des experts indiquent en effet que l'expert Direction F est dévalué plus rapidement que l'expert Planification des autres groupes et donc l'octant F ne peut pas être occupé dans ces essais (Figure 6.3 c).

Le groupe T-par présente l'effet d'extinction, pas de gradient. Le groupe T-par ne présente pas d'occultation pendant les essais d'extinction, car les deux experts ancrés sur les amers B et F ont des taux de sélection plus élevés (Figures 6.2 c et d droit, voir annexe D.1). Les essais d'extinction ont eu l'effet escompté sur lui, mais ont aussi causé aussi son manque de gradient dans l'octant B. On note effectivement un taux de sélection important de l'expert Taxon B dans l'octant B (Figure D.2).

6.1.2.4 Pourquoi tous les groupes présenteraient une augmentation d'occupation de l'octant B à 180°

Nous avançons ici une hypothèse qu'il faudrait néanmoins étayer plus profondément. L'exemple de la carte de planification, dont on rappelle qu'elle se construit avec un seul amer, présente des zones beaucoup moins cartographiées que d'autres, dont la zone 180° (Figure 6.5). Dans ces zones, le module de sélection accorderait donc à l'expert Direction B plus de valeur et entraînerait plus souvent les animats vers l'amer B.

6.1.3 Discussion partielle

Les groupes avec les seuls experts procéduraux ne reproduisent que partiellement cette expérience. Dans le groupe avec les seuls experts directionnels, les effets de gradient associatifs lié à l'amer proximal et l'absence de gradient lié à l'amer distal peuvent y être observés, mais les effets d'atténuation de l'occultation de l'amer distal recherché par les auteurs ne peuvent pas être reproduits. Ce groupe dévalue de façon trop importante cet amer par rapport à un autre groupe constitué d'un expert de lieu. Dans le groupe avec les seuls experts Taxon, l'amer distal est trop dévalué et le gradient d'occupation dans l'octant de l'amer proximal n'est plus visible.

Celui qui semble le mieux produire le comportement des animaux réels doit comporter un expert Direction et un expert Planification s'ancrant chacun sur un amer respectif (groupe DP-sep, entrées séparées). La pertinence de l'amer distal y est en effet mieux apprise que dans les autres groupes. L'effet d'occultation de l'amer distal par le proximal y est donc moins fort, grâce à un apprentissage supplémentaire pendant les sessions d'extinctions, ce que les auteurs avaient cherché à faire.

L'effet de gradient associatif peut donc s'expliquer avec le seul concours d'experts directionnels procéduraux, mais l'existence d'une stratégie de planification semble ici essentielle pour expliquer l'atténuation de l'occultation de l'amer distal par les auteurs. Les expérimentateurs supposent au contraire que tous les effets observés ne sont dus qu'à une compétition entre les deux amers, appris tous les deux par une règle associative. Nos analyses proposent en effet une telle explication, mais associée à la nécessité d'utilisation d'une représentation spatiale intégrant l'amer distal.

Une remarque supplémentaire est que le gradient de valeur associative lié à la distance entre l'amer proximal et distal doit être distingué en deux phases. La première concerne le premier déplacement jusqu'à l'octant de l'amer déplacé, qui s'effectue très rapidement pour tous les groupes et sans gradient, majoritairement à l'aide de l'expert de guidage ancré sur l'amer proximal. Cette constatation n'avait pas été signalée par les auteurs. Une deuxième phase concerne spécifiquement l'effet de gradient par une recherche de la plateforme absente, qui produit l'effet par des interactions compétitives, augmentant avec la distance entre les amers, entre l'expert de guidage et l'expert de lieu.

6.2 Effet de blocage d'amers distaux par un amer proximal dans l'expérience de Roberts et Pearce (1999)

L'objectif de cette expérience était de démontrer :

- (i) qu'un apprentissage de plusieurs amers distaux peut être bloqué par l'apprentissage *préalable* d'un amer proximal situé sur le but - donc que l'apprentissage spatial suit une loi associative.
- (ii) que plus l'amer proximal est présent, plus le blocage est important.

Si ces résultats sont avérés, sur la base de nos résultats précédents, notre hypothèse est que

l'expert de guidage, suffisamment entraîné lorsque l'expert Planification ne peut pas être efficace, va prendre le contrôle sur cet expert de lieu, même si celui-ci a accès ensuite à un apprentissage. Le module de sélection n'aura en effet pas eu le temps d'apprendre l'efficacité de ce dernier expert.

6.2.1 Protocole et résultats observés

L'expérience n° 3 de Roberts et Pearce (1999) se situe dans une piscine de Morris de 2 m de diamètre, dans laquelle est placée une plateforme invisible de 10 cm de diamètre. Un amer de 13 cm de diamètre est attaché à cette plateforme. Les animaux ont aussi accès à plusieurs amers distaux. Quatre groupes d'animaux prennent part à l'expérience, consistant en 15 sessions de 4 essais (Figure 6.6 a).

Pour les groupes Session, dans la Phase 1 de 12 sessions, la plateforme est déplacée régulièrement toutes les sessions. Dans Phase 2 de 3 sessions, la plateforme et l'amer associé restent à la même place (à 50 cm du mur Nord-Ouest). Après le 4^{ème} essai de la 3^{ème} session, un essai de test est administré (Phase 3), durant lequel la plateforme et l'amer sont retirés. A ce moment-là le temps que passent les rats dans le quadrant où se trouvait la plateforme est relevé : si les animaux n'ont rien appris de l'emplacement du but par rapport aux indices distaux, ils devraient ne montrer aucune préférence pour ce quadrant.

Un groupe Session-Same garde le même amer A pendant toute l'expérience. Pour un groupe Session-Diff, les caractéristiques physiques de l'amer sont changées pendant la deuxième Phase (amer B de 10 cm de diamètre, de forme cylindrique et de couleur différente). L'hypothèse est que les deux groupes ne vont pas montrer de blocage de l'apprentissage des amers distaux et que le changement d'amer proximal va réduire encore plus l'effet de blocage.

Pour les groupes Trial-Same et Trial-Diff, le protocole de la Phase 1 de 12 sessions est le même que pour les groupes Session-Same et Session-Diff, mais au lieu de déplacer la plateforme et l'amer A à chaque session pendant la première Phase, ils sont déplacés à chaque essai. Ces groupes ne vont donc pas s'ancrer sur les amers distaux, contrairement aux groupes Session. La Phase 2 de 3 sessions est identique aux groupes Session, c'est-à-dire que les groupes Trial peuvent apprendre la position de la plateforme avec les amers distaux et proximaux (A ou B) et ont un essai de test à la Phase 3, dans lequel l'amer et la plateforme sont retirés. Le groupe Trial-Same, qui garde le même amer A pendant toute l'expérience, a l'occasion d'apprendre plus fortement la position de la plateforme par rapport à cet amer que le groupe Trial-Diff, qui a eu à réapprendre un autre indice B. L'hypothèse est, que par cette opportunité d'avoir appris encore plus la position de la plateforme par l'amer A, seul Trial-Same devrait présenter un blocage des amers distaux.

La Figure 6.6 b montre les courbes d'apprentissage observées, mais seulement pendant la Phase 2. On remarque que si les groupes Session-Same et Trial-Same ne montrent aucun apprentissage, les groupes Session-Diff et Trial-Diff réapprennent à retrouver le but, alors marqué par un amer différent de la première Phase. La Figure 6.6 c indique en effet que seul le groupe Trial-Same

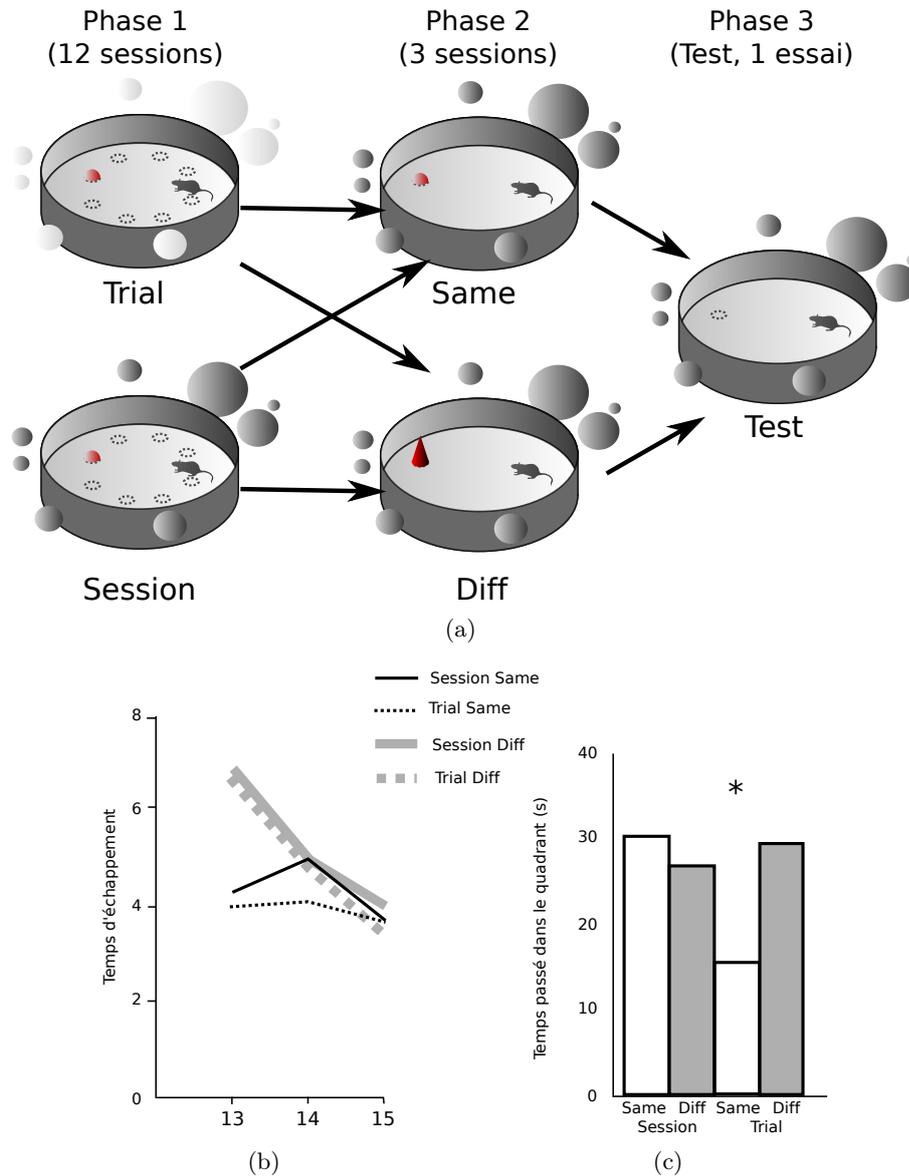


Figure 6.6 – **Expérience de Roberts et Pearce (1999)**. (a) Protocole : la Phase 1 est divisée en deux groupes : le groupe Session où la plateforme est déplacée toutes les sessions (4 essais) et le groupe Trial où la plateforme est déplacée tous les essais. La Phase 2 est également divisée en deux groupes : le groupe Same, où l'amer rattaché à la plateforme est le même qu'en Phase 1 et le groupe Diff avec un amer différent. Enfin l'essai de test est réalisé en l'absence d'amer proximal. (b-c) Résultats originaux de l'expérience de Roberts et Pearce (1999). (b) Apprentissage des quatre groupes, pendant la Phase 2 (sessions 13 à 15). (c) Temps passé par les différents groupes dans le quadrant du but pendant l'essai de test.

ne va pas plus qu'au hasard dans le quadrant de la plateforme lors de l'essai de test, les autres groupes ayant tous une nette préférence pour ce quadrant. Conformément à leurs hypothèses, les auteurs en concluent :

- (i) qu'un entraînement préalable sur des amers distaux et proximaux influence l'effet de blocage, puisqu'il y a des différences entre les groupes Session et ce groupe Trial-Same
- (ii) qu'un amer proximal préalablement appris influence l'effet de blocage, puisqu'il y a des différences entre les groupes Trial-Same et Trial-Diff.

6.2.2 Protocole et résultats simulés

Nous reproduisons le même environnement que l'original en termes de distances et de tailles (but, plateforme, amer proximal A). Les amers distaux sont au nombre de 8, de tailles diverses et pseudo-aléatoirement répartis autour de la piscine simulée.

Comme le but est ici directement indicé, les groupes simulés comprendront aussi un expert Taxon. Comme le but est fixe pendant la Phase 2, on inclut aussi un expert Local. Les groupes suivant sont donc simulés :

- groupe TP-sep, dans lequel l'expert Taxon s'ancre sur l'amer A (ou l'amer B) et l'expert Planification sur les 8 amers distaux
- groupe TP-par, dans lequel l'expert Taxon et l'expert Planification s'ancrent sur l'amer A (ou l'amer B) et les 8 amers distaux.
- groupe T, dans lequel des experts Taxon s'ancrent sur l'amer A ou l'amer B (Taxon A ou Taxon B) et les 8 amers distaux. Ce groupe indiquera la nécessité ou non d'une construction d'une carte cognitive.
- groupes DP-sep, DP-par et D-par seront aussi simulés, avec les mêmes caractéristiques que précédemment, l'expert Taxon étant remplacé par un expert Direction (soit Direction A ou Direction B).
- enfin un groupe DL-sep, avec les mêmes caractéristiques que précédemment, l'expert Planification étant remplacé par un expert Local. Les résultats du groupe DL-par, similaires à ceux du groupe DL-sep, sont présentés en annexe D.

Nous estimons la reproduction de la tâche en calculant le pourcentage de pas de temps passé par les animats dans le quadrant correct lors de l'essai de test. Nous évaluons le choix des amers distaux par les experts Direction en calculant leur valeur de sélection pendant l'entraînement et calculons les taux de sélection des experts lors de l'essai de test, durant la trajectoire *5 pas de temps avant* que l'animat ne rentre dans le quadrant correct, et ce afin de connaître l'expert responsable d'un tel comportement.

6.2.2.1 Le groupe DP-sep est le meilleur candidat pour reproduire le comportement des animaux

Phase 1 d'entraînement (sessions 1 à 12). Les Figures 6.7 a, c et 6.8a montrent l'évolution des temps d'échappement pour les groupes respectivement TP, DL et DP. En début d'entraînement, les différences entre les animats Session et Trial ne s'observent que chez les groupes TP-sep, DP-sep et DP-par, dans lesquels l'expert Planification peut construire une carte en intégrant les

amers distaux. En revanche, les groupes TP-par, avec un seul type d'expert (T ou D), ou avec l'expert Local ne présentent pas cette différence, qui est signe qu'un certain apprentissage n'a pas pu avoir lieu pendant une seule session.

Nous ne pouvons pas comparer ces courbes d'apprentissage avec celles des animaux pendant l'entraînement, celles mentionnées dans l'article ne correspondant qu'à la Phase 2. La prédiction du modèle concernant la différence de performances des groupes Trial et Session en début d'expérience pourrait cependant aisément être testée.

Phase 2 (avec amers distaux et proximal A ou B ; sessions 13 à 15). Tous les groupes présentent la même différence de performance entre animats Same et Diff (Figure 6.7 a, c et 6.8 a). Chez les animaux de l'expérience originale, on constate que ces performances deviennent les mêmes en session 15 (Figure 6.6 b). Sans obtenir des performances identiques en fin de Phase 2, les groupes d'animats qui correspondent le mieux à cette observation sont les groupes TP-sep et DP-sep.

Essai de test (avec amers distaux, sans amer proximal et sans plateforme). Les graphes de la Figure 6.7 b, d et 6.8 b montrant le temps d'occupation dans le quadrant du but pendant l'essai test, indiquent clairement que l'effet de blocage des amers distaux par les amers proximaux constaté dans les seuls animaux Trial-Same ne se révèle que dans les groupes où les experts Direction et Planification sont en présence, soit DP-sep et DP-par. Les autres groupes présentent différentes formes de blocage ou d'absence de blocage pour les différents animats, mais pas dans la configuration correspondant à l'expérience originale.

Le groupe DP-sep (Figure 6.8 b, gauche) semblerait présenter des similitudes plus grandes avec les groupes d'animaux car, contrairement au groupe DP-par, il ne présente pas de différences significatives entre temps d'occupation des animats Session-Same, Session-Diff et Trial-Diff.

Compte tenu de ces similarités, nous allons analyser en priorité pourquoi l'effet de blocage produit dans l'expérience originale a été obtenu par le groupe DP-sep.

6.2.2.2 Groupe D-sep. Phase 1 : différence de sélection d'experts chez les animats Session et Trial (Figure 6.10)

L'expert Planification a été plus sélectionné en début d'expérience chez les animats Session. Les taux de sélection dans les groupes DP-sep montrent que l'expert Planification est sélectionné plus fréquemment dans tous les animats Session pendant le début de l'entraînement (Figure 6.10 a, c). C'est sa contribution qui a pu améliorer l'apprentissage des animats Session : la Figure 6.11 montre en effet que la performance des animats Session s'améliore considérablement au cours d'une session, comparée à celle du groupe Trial – où les amers distaux sont difficilement accessibles. Nous reviendrons sur l'explication de cette amélioration dans le chapitre suivant (cf. reproduction de l'expérience de Pearce et al., 1998). Ces résultats signifient que la position de la plateforme par une stratégie de lieu peut être apprise rapidement par ces

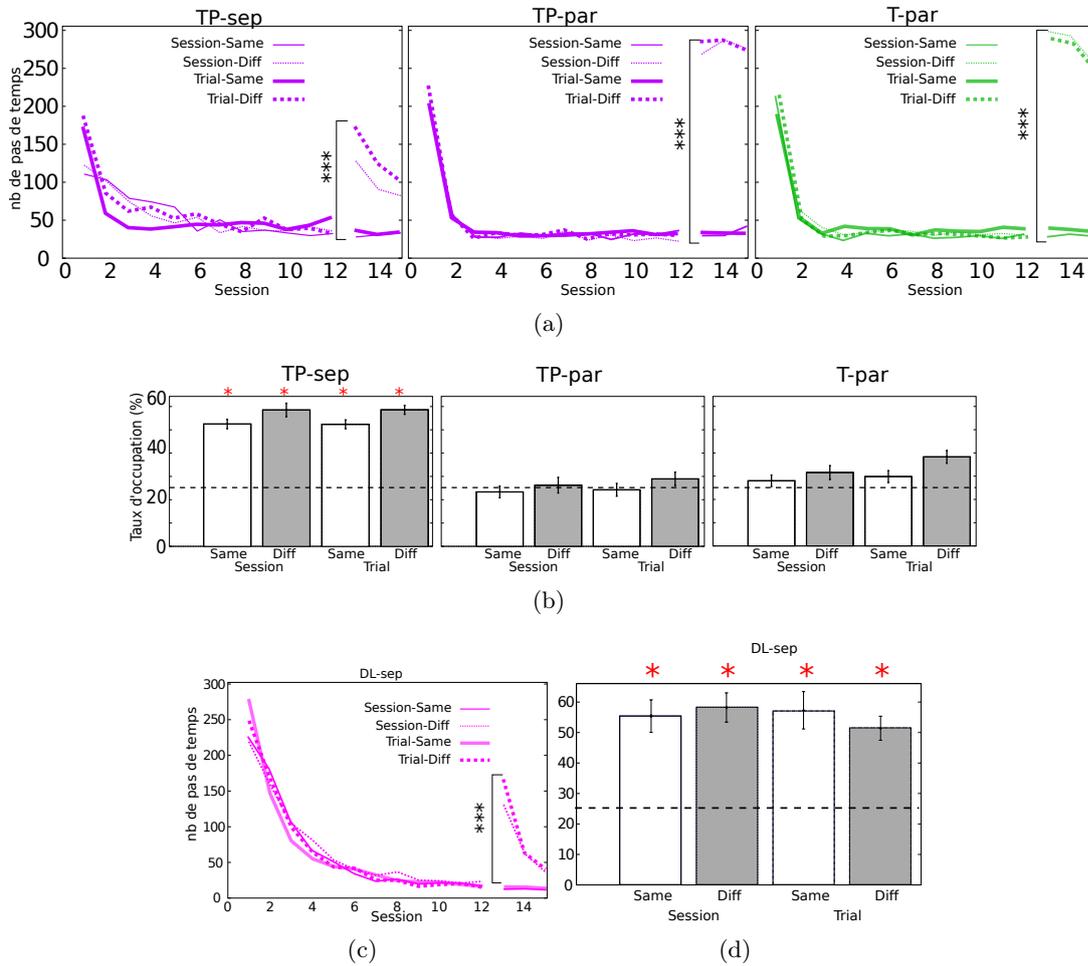


Figure 6.7 – **Simulation de l'expérience de Roberts et Pearce (1999).**(a, c) Courbes d'apprentissage des groupes avec (a) expert Taxon (TP-sep, -par, T), (c) expert Local (DL-sep). (b, d) Temps passé par les différents groupes dans le quadrant où se trouvait le but avant l'essai de test.

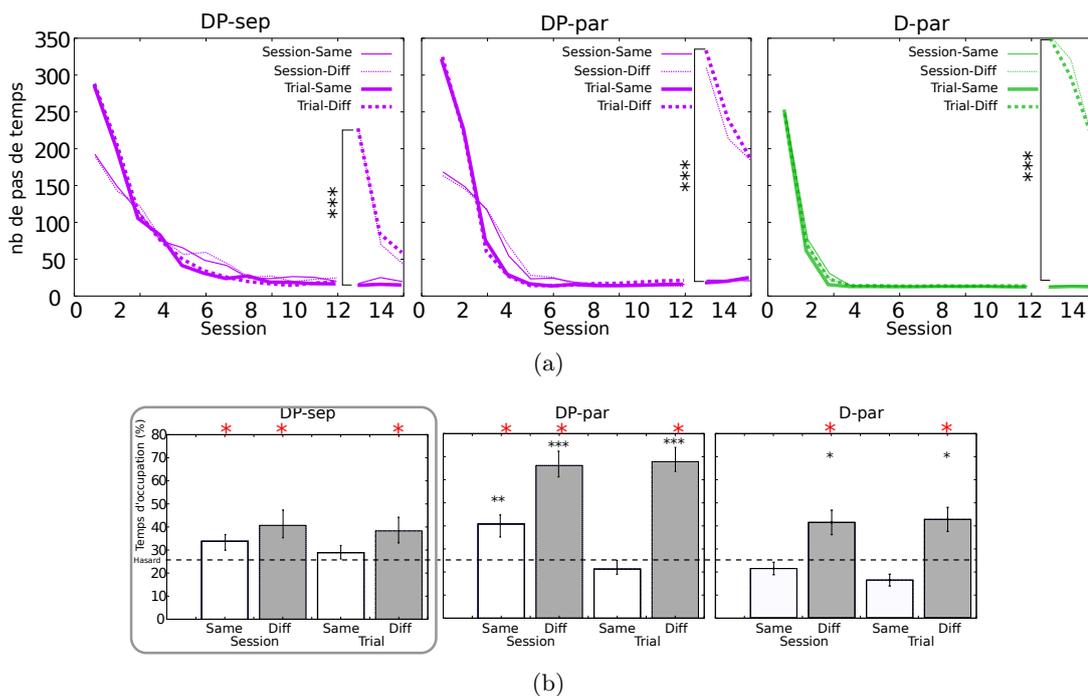


Figure 6.8 – **Simulation de l'expérience de Roberts et Pearce (1999).**(a) Courbes d'apprentissage des groupes avec experts Direction (DP-sep, -par et D). (b) Temps passé par les différents groupes dans le quadrant où se trouvait le but avant l'essai de test. Le groupe DP-sep est conforme aux résultats observés, du point de vue de l'apprentissage comme du temps passé dans le quadrant.

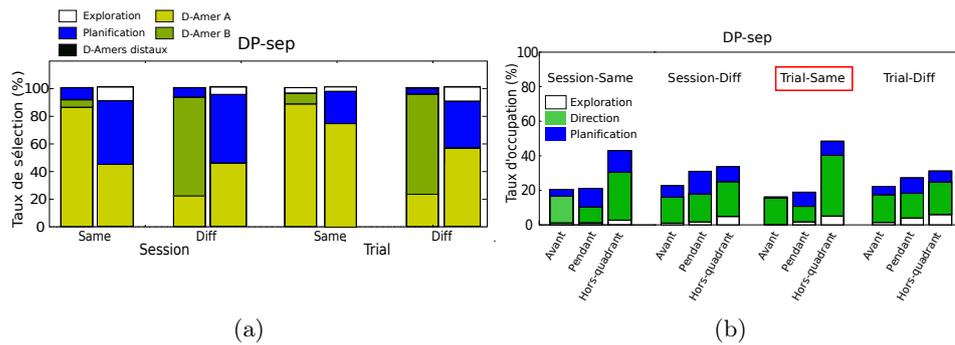


Figure 6.9 – **Simulation de l'expérience de Roberts et Pearce (1999)**. (a) Sélection des différents experts du groupe DP-sep *juste avant* l'essai de test (première barre de chaque groupe) et *pendant* l'essai de test (deuxième barre). (b) Détail des taux de sélection pendant les 3 principales étapes de la trajectoire pendant l'essai de test, *avant de rejoindre pour la première fois* le quadrant de la plateforme (première barre), puis *ensuite*, lorsqu'ils occupent ce quadrant (deuxième barre) ou un autre (troisième barre). Les taux sont pondérés par les temps effectifs d'occupation. Les animaux Trial-Same, seuls à exprimer un blocage, sont entourés en rouge.

animats (qu'ils partagent ou non tous les types d'amers, puisque le groupe DP-par présente le mêmes caractéristiques).

Chez les animats Trial, on constate une chute brusque de la sélection de l'expert de lieu en début d'expérience (Figure 6.10 b, d), le module de sélection l'ayant dévalué au profit de l'amer proximal présent.

L'expert Direction A prend plus rapidement le contrôle chez les animats Trial. L'expert Direction A acquiert immédiatement un contrôle sur le comportement des animats Trial (Figure 6.10 b, d). Cet expert prend également le contrôle dans les animats Session, mais après une sélection conjointe avec l'expert de lieu (Figure 6.10 a, c).

6.2.2.3 Phase 2 : ré-apprentissage de l'amer B chez les animats Diff

Les performances des animats Same sont significativement meilleures que celles des animats Diff. On rappelle que, dans cette Phase, la plateforme et l'amer restent à la même place pour tous les groupes, mais que les caractéristiques de l'amer proximal soient modifiées pour les animats Diff (amer B). Les animats Same montrent le même niveau d'apprentissage que dans la Phase 1, alors que les animats Diff présentent une augmentation considérable de leurs temps d'échappement (Figure 6.8 a gauche) : ils ont à réapprendre la pertinence d'un autre amer. Il n'y a pas de différence de performance entre les animats Session-Diff et Trial-Diff (des amers distaux sont présents dans les deux cas).

L'expert de lieu n'est plus sélectionné chez les animats Same. La position du but avec l'amer A pendant la Phase 2 est donc apprise majoritairement par l'expert de guidage (Figure 6.10 a, b).

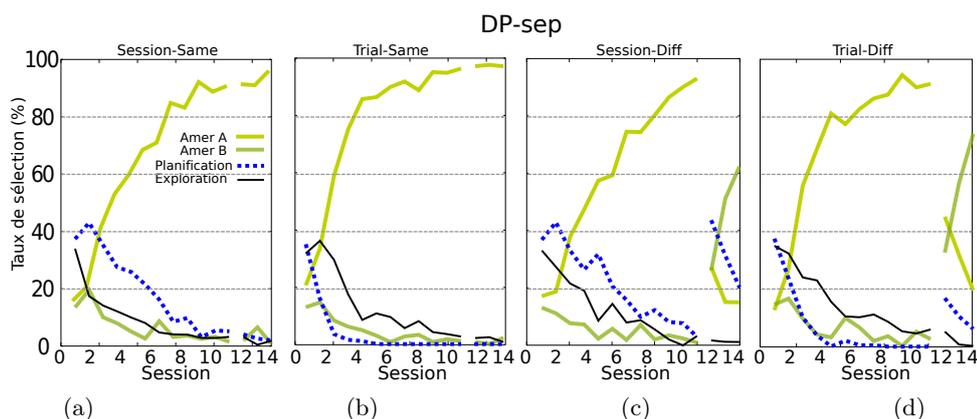


Figure 6.10 – **Simulation de l'expérience de Roberts et Pearce (1999)**. Taux de sélection des experts correspondants au groupe DP-sep. (a) groupe Session-Same, (b) groupe Trial-Same, (c) groupe Sessin-Diff, (d) groupe Trial-Diff.

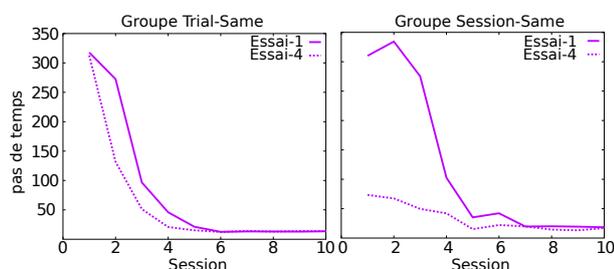


Figure 6.11 – **Simulation de l'expérience de Roberts et Pearce (1999)**. Comparaison des temps d'échappements dans les premiers et quatrièmes essais d'une Session dans les groupes DP-sep Session et Trial.

Les deux types d'experts sont mis à contribution chez les animaux Diff. L'expert Direction réapprend en effet la pertinence d'un autre amer (B) avec le contrôle majoritaire de l'expert Direction B, mais accompagné d'une augmentation de la sélection de l'expert de lieu – plus importante chez les animaux Session-Diff, ce qui est attendu, puisque l'expert de lieu y a été moins dévalué dans la Phase 1 (Figure 6.10 c, d).

6.2.2.4 Essai test : effets de blocage des amers distaux

Pourquoi il y a un effet de blocage chez les animaux Trial-Same. On rappelle que l'expert Direction A prend majoritairement le contrôle. Depuis le début de l'expérience, il est beaucoup plus sélectionné que l'expert Planification - puisque les amers distaux ne sont pas fiables. La Figure 6.9 a indique le détail de sélections d'experts avant (premières barres) et pendant l'essai test (deuxièmes barres). Le groupe Trial-Same est le troisième de cette Figure. On constate que l'expert Direction A, avant le test, est donc toujours sélectionné de façon

privé au détriment de l'expert Planification. Sans le seul amer qui pourrait les aider dans leur orientation, les animats ne peuvent pas réussir à rejoindre la plateforme.

Cette explication n'est cependant pas exacte, car la Figure 6.9 a montre que l'expert Planification est tout de même sélectionné pendant le test, même si juste avant il ne l'était pas.

En détaillant la sélection d'experts pendant cet essai de test avant de rejoindre pour la première fois le quadrant de la plateforme (première barre) puis ensuite, lorsqu'ils occupent ce quadrant (deuxième barre) ou un autre (troisième barre) (Figure 6.9 b), on constate que, contrairement aux autres animats, les Trial-Same continuent à exclusivement sélectionner l'expert Direction A *avant* de rejoindre le quadrant (première barre), mais qu'ils recrutent l'expert Planification pour rechercher la plateforme ensuite. Il n'est pas très efficace et les animats passent un temps important hors du quadrant (troisième barre vs deuxième), ce qui renforce l'effet de blocage.

En résumé, les animats Trial-Same n'ont bloqué l'utilisation des amers distaux (et donc la sélection de l'expert Planification) que lorsqu'ils essayent de rejoindre pour la première fois la plateforme, tentant de s'ancrer sur un amer proximal qui est absent. Le quadrant de la plateforme une fois atteint (par hasard), ils recrutent les amers distaux pour rechercher ensuite la plateforme absente. C'est une distinction (différence de sélection entre la première atteinte de la zone de la plateforme et ensuite) que nous avons noté pour l'effet de gradient de Rodrigo et al (2006), mais ici l'effet de blocage s'effectuerait dans la première Phase. Cette possibilité n'a pas été évoquée par les auteurs de l'expérience originale.

Pourquoi il y n'a pas d'effet de blocage chez les autres animats (Session-Same, Session-Diff et Trial-Diff) de ce groupe DP-sep. Le groupe Session-Same possède à la fois un expert Direction et un expert Planification qui ont pu apprendre la position du but pendant l'entraînement. L'expert Direction A est beaucoup plus efficace et continue à être sélectionné exclusivement pendant la Phase 2. Pendant le test, les animats sélectionnent l'expert de lieu (qui a pu configurer correctement sa carte pendant les sessions), même si l'expert de guidage est aussi recruté (Figure 6.9, a) et ce, dès la première recherche de la plateforme (Figure 6.9 b premier groupe).

Puisqu'ils subissent le même entraînement, les experts Direction et Planification des animats Session-Diff deviennent aussi efficaces que ceux des animats Session-Same ; l'expert Direction des animats Trial-Diff devient aussi efficace - et l'expert Planification aussi peu efficace - que ceux des animats Trial-Same. Le changement d'amer dans la Phase 2, donc le changement d'expert Direction, dévalue l'efficacité de l'expert Direction précédent dans le module de sélection, et les deux experts de guidage et de lieu se trouvent dans la même situation qu'au tout début de l'entraînement. Ils coopèrent donc pour augmenter rapidement la performance des animats, comme au début de l'expérience.

Avec la contribution de l'expert Planification pendant la Phase 2, les animats Diff ont suffisamment appris à rejoindre le quadrant de la plateforme pour l'occuper pendant l'essai de test (Figure 6.9 a. Première barre : augmentation de la sélection de Direction B ; deuxième barre : recrutement de l'expert Planification) et ce, dès la première recherche de la plateforme (Figure 6.9 b, première barre, deuxième et troisième groupes). Ils passent un temps plus important que les

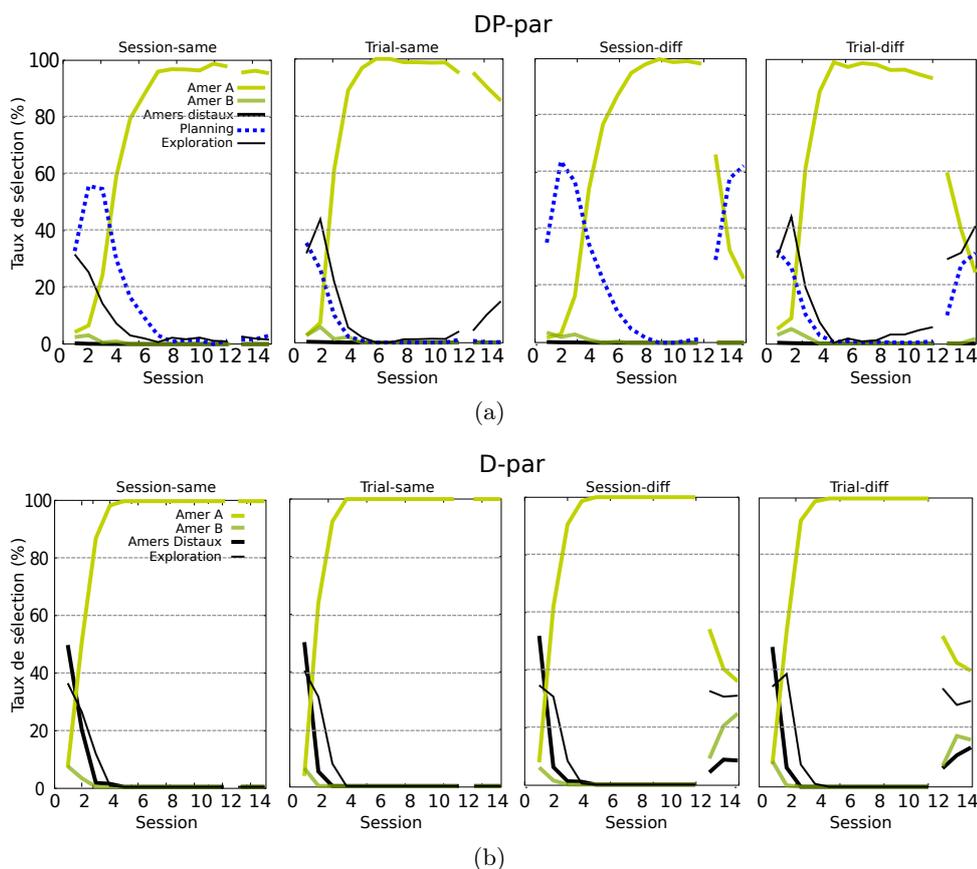


Figure 6.12 – **Simulation de l'expérience de Roberts et Pearce (1999)**. (a) Taux de sélection des experts correspondants aux groupes DP-par ; et (b) D-par.

Same dans le quadrant de la plateforme, par suite d'un recrutement plus important de l'expert Planification, causé par une remise en cause des sélections dans le module correspondant.

6.2.2.5 Effet ou absence de blocage dans les autres groupes (DP-par et D)

Dans le groupe DP-par, l'effet et l'absence de blocage sont renforcés. Les explications concernant ce groupe sont les mêmes que pour le groupe DP-sep. L'expert Planification a été bien sélectionné chez les groupes Session-Same et les groupes Diff. Il n'est en compétition qu'avec l'expert Direction A, qui a gagné une valeur associative plus grande car il a été en compétition avec 8 autres experts (Figure 6.12 a) : juste avant le test, l'expert B n'est même pas sélectionné dans les groupes Diff (Figure 6.13 a, troisième groupe). L'expert de lieu peut donc entraîner les animats vers le quadrant de la plateforme.

Chez les animats Trial-Same de ce groupe, leur utilisation de l'expert de lieu est minime pendant l'essai test, pendant la première recherche de la plateforme et le reste de la trajectoire (Figure

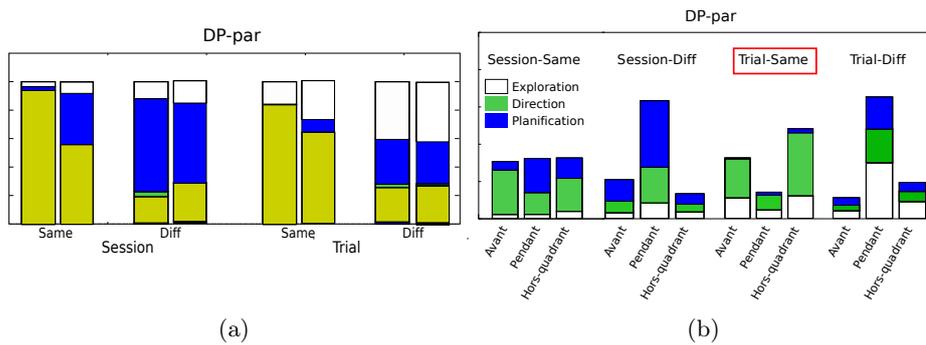


Figure 6.13 – **Simulation de l'expérience de Roberts et Pearce (1999)**. (a) Sélection des différents experts du groupe DP-par *juste avant* l'essai de test (première barre de chaque groupe) et *pendant* l'essai de test (deuxième barre). (b) Détail des taux de sélection pendant les 3 principales étapes de la trajectoire pendant l'essai de test, *avant de rejoindre pour la première fois* le quadrant de la plateforme (première barre), puis *ensuite*, lorsqu'ils occupent ce quadrant (deuxième barre) ou un autre (troisième barre). Les taux sont pondérés par les temps effectifs d'occupation. Les animats Trial-Same, seuls à exprimer un blocage, sont entourés en rouge.

6.13 b, droit, troisième groupe). L'expert Direction A prend en effet un contrôle encore plus important que dans le groupe DP-sep, ce qui renforce l'effet de blocage observé.

Dans le groupe D, seuls les deux groupes d'animats Same présentent l'effet de blocage. Les 9 experts Direction des animats Same ont été sélectionnés, puis les 8 distaux ont été dévalués par le module de sélection pendant les Phases 1 et 2. L'expert Direction A a majoritairement pris le contrôle (Figure 6.12 b), donc aussi pendant l'essai de test. En revanche, chez les animats Diff, le ré-apprentissage de l'amer B pendant la Phase 2 a brusquement dévalué l'expert Direction A, recrutant à nouveau l'expert B et certains experts distaux (6.12 b), qui deviennent suffisamment efficaces (avec de l'exploration) pour conduire les animats dans le quadrant de la plateforme pendant l'essai de test).

6.2.2.6 Cas de la présence d'un expert Taxon (résultats détaillés dans l'annexe D.2).

On peut se demander en effet pourquoi un groupe avec expert Taxon égocentré ne reproduit pas beaucoup mieux les effets observés, étant donné que l'amer proximal est situé sur le but.

Chez tous les animats du groupe TP-sep, on n'observe aucun blocage (Figure 6.7 b, gauche). L'expert Taxon A prend très rapidement le contrôle sur l'expert de lieu, mais celui-ci a pu être sélectionné en début d'expérience, beaucoup plus dans les groupes Session. On observe qu'avant l'essai de test, les experts Taxon A et B sont plus sélectionnés chez les animats Same et moins chez les animats Diff que dans leurs équivalents du groupe DP (Figure D.3 en annexe). Cela dénoterait une plus forte compétition entre eux deux – et avec l'autre expert – ce qui pourrait

laisser la possibilité au module de sélection de favoriser l'expert de lieu pendant l'essai test – ayant appris suffisamment pendant le début de l'entraînement chez les animats Same, et pendant la Phase 2 chez les animats Diff (Figure D.4 en annexe).

Tous les autres animats (TP-par et T-par) montrent un effet de blocage (Figure 6.7 b, milieu et droit). Pendant la Phase 1, la forte dévaluation des 8 experts Taxon (distaux) et de l'expert Planification empêcherait un ré-apprentissage efficace des amers distaux.

En résumé, notre proposition est que les effets de blocage non conformes à ceux des animaux dans les groupes avec experts Taxon proviendraient d'une compétition plus forte entre experts que les groupes avec experts Direction. C'est ce que nous avons mis en évidence dans la section "Amers seuls, but visible", où le groupe TP-sep montrait une moins bonne performance que le groupe T – alors que les groupes DP-sep et D ne montraient aucune différence (même si, dans ces situations, l'expert P n'avait qu'une carte "odométrique").

6.2.2.7 Cas de la présence d'un expert Local (résultats détaillés dans l'annexe D.2)

Un but fixe avantage l'expert Local par rapport à l'expert Planification. En effet, il peut développer une carte précise des différents emplacements du but (Figure 8.2 dans la Discussion).

Tous les groupes avantagent l'expert Direction pendant l'entraînement, mais le groupe Trial-Same augmentent la sélection de l'expert Local, sélectionné de façon privilégiée à cause de sa précision, pendant la Phase 2. Les animats de ce groupe réussissent donc à rejoindre le quadrant de la plateforme aussi bien que les autres groupes.

6.2.3 Discussion partielle

Les résultats indiquent qu'un effet de blocage des amers distaux par un amer proximal tel que celui démontré par Roberts et Pearce (1999) (avec absence de blocage dans certains groupes) ne peut pas s'obtenir avec les seuls experts procéduraux, égocentrés ou allocentrés. Chez les premiers, l'amer proximal appris en même temps qu'avec les amers distaux bloque les amers distaux en toute circonstance, chez les seconds le lent apprentissage avec amers distaux provoque un blocage dans un groupe qui ne devrait pas en exprimer. Un groupe Trial-Same avec expert Local n'exprime pas l'effet de blocage.

Un groupe constitué d'un expert Direction et d'un expert Planification, focalisant respectivement sur l'amer proximal et les amers distaux (DP-sep), représenterait le mieux le groupe d'animaux. La précision d'un expert Direction bien entraîné, mais aussi la flexibilité de l'expert de lieu, seraient cause de la présence et de l'absence de blocage dans certains groupes. L'hypothèse de blocage des amers distaux par un amer proximal ne serait donc pas incompatible avec la présence d'un expert utilisant une carte topologique.

L'effet de blocage des amers distaux se révélerait mieux dans la première approche du quadrant de la plateforme absente, l'expert de lieu pouvant être réutilisé ensuite pour rechercher le but absent. Les auteurs n'avaient pas indiqué cet effet, qu'il serait possible de confirmer en évaluant,

chez un groupe Trial-Same, l'efficacité de la navigation avant et après avoir atteint pour la première fois le quadrant du but.

Dans leur discussion, les auteurs abordent cependant un certain nombre d'éventualités pour expliquer leurs observations. Ils supposent par exemple que l'absence de blocage des amers distaux chez les animaux Session-Same pourrait s'expliquer par le manque d'entraînement pour l'apprentissage de l'amer proximal. Or, l'amer A est au contraire très bien appris (environ dès la quatrième session), mais l'expert de lieu est aussi pertinent, lui permettant de retrouver le quadrant de la plateforme en l'absence de l'amer proximal.

Les auteurs suggèrent aussi que le blocage a été induit par un processus attentionnel, forçant les animaux Trial-Same de se focaliser sur l'amer proximal. Mais on constate dans les simulations que tous les animaux, présentant ou non un effet de blocage, focalisent sur l'amer proximal en fin d'entraînement. Ce pourrait donc plutôt être l'opportunité d'avoir ou non pu utiliser une représentation spatiale fiable qui explique le blocage.

Ils rejettent d'ailleurs la possibilité d'utilisation d'une telle intégration d'amers comme étant la seule stratégie mise en œuvre. En effet, la situation dans laquelle les experts intègrent tous les amers ne reproduit pas exactement le pattern de tous les animaux testés – bien qu'il en soit proche. Des interactions de stratégies focalisant chacune sur un type d'amer semblent donc plus plausibles.

Ils supposent aussi une autre alternative, une utilisation d'amers individuels identifiés par une stratégie "heading vector", qui pourrait avoir positionné la plateforme en apprenant sa distance et sa direction par rapport à ces amers. Notre modèle montre que cette supposition est possible - l'expert Direction se révélant efficace pour rejoindre le quadrant du but – mais la seule utilisation de cette stratégie de guidage ne peut pas expliquer l'absence de blocage chez les animaux Session-Same.

Les auteurs évoquent finalement la possibilité de l'utilisation de plusieurs stratégies, sans pouvoir vérifier cette hypothèse, supposant qu'une stratégie de lieu ait pu être utilisée sur la base d'une intégration d'amers distaux et qu'une stratégie "heading-vector" ait pu être utilisée sur la base de l'amer proximal – ajoutant que dans ce cas, cela impliquerait que ces stratégies ne se développeraient pas indépendamment les unes des autres. C'est cette explication que nous avons étayée ici.

6.3 Absence d'occultation d'amers proximaux par les amers distaux (McGregor et al, 2004a)

L'expérience de McGregor et al (2004a), que nous avons partiellement reproduite dans la section 5.1.3, vise à vérifier si les relations entre amers proximaux et distaux suivent une loi associative, en particulier si l'apprentissage *simultané* des amers distaux peut occulter celui des amers proximaux. Les auteurs trouvent au contraire que les deux apprentissages s'effectuent en

parallèle et que les animaux qui ont appris l'emplacement du but en fonction des amers distaux peuvent aussi retrouver le but en fonction des seuls amers proximaux.

Si ces résultats sont avérés, notre hypothèse est que les experts Direction et Planification peuvent se focaliser sur chaque type d'amers et que les deux stratégies de guidage et de lieu seront suffisamment apprises pour pouvoir s'exprimer si besoin est. Les groupes avec expert Taxon ne seront pas simulés, compte tenu de leur résultats dans les groupes Rotate (voir section 5.1.3.2).

6.3.1 Protocole et résultats observés

Comme pour le groupe Rotate (cf. Chapitre 6), dont nous allons rappeler les résultats ici, le groupe Stable de l'expérience 2 doit trouver une plateforme cachée de 2 m de diamètre, indiquée par deux amers proximaux A et B. Contrairement au groupe Rotate, ce groupe a accès aux indices distaux de la pièce et peut donc aussi apprendre la position de la plateforme par ces amers. L'entraînement se déroule sur 15 sessions de 4 essais (Figure 6.14, a gauche). Les essais 2 et 4 des sessions 12 et 14 sont des essais de test : ceux-ci se déroulent avec un seul amer (différent selon les animaux, A ou B) dans la session 12 (Figure 6.14, a milieu) et les deux dans la session 14 (Figure 6.14, a droite). Dans chacune de ces sessions, la plateforme reste au même endroit "correct" dans un essai et est déplacé dans un endroit "opposé" au segment constitué par les amers dans l'autre essai. L'ordre de ces essais est variable selon les individus (correct/opposé ou opposé/correct). Les rideaux sont fermés dans la session test 14, permettant ainsi d'évaluer la performance des rats sans les indices distaux (Figure 6.14, a droit).

Les auteurs ont relevé les performances des rats dans les deux Phases (Figure 6.14 Rotate : b, Stable : c) et ont ensuite comparé le temps nécessaire aux animaux pour rejoindre la plateforme dans chacun de ces essais (Figure 6.14 d) : Dans la session test 12 (où un seul amer proximal est présent), les animaux du groupe Rotate rejoignent la plateforme au hasard dans les essais correct et opposé, tandis que ceux du groupe Stable présentent des différences significatives dans l'essai correct, avec de meilleures performances que dans l'essai opposé. Dans la session test 14 (où deux amers proximaux sont présents et les distaux absents), les animaux des deux groupes sont nettement plus performants dans l'essai correct que dans l'essai opposé.

Les résultats de la session 12 démontrent que, contrairement au groupe Rotate, les animaux du groupe Stable ont appris la position de la plateforme par rapport aux indices distaux. Cet apprentissage n'a cependant pas occulté l'apprentissage des amers proximaux puisque le groupe Stable est capable de retrouver l'emplacement de plateforme seulement par sa position relative à ces deux amers, comme le montrent les résultats de la session 14.

Enfin les auteurs ont noté que, pendant l'entraînement, les animaux du groupe Stable prenaient contact avec un seul amer proximal (voire aucun) et non deux comme ceux du groupe Rotate, avant de rejoindre le but. Mais dès que le rideau était tiré, donc en l'absence d'amers distaux, les animaux exprimaient à nouveau ce contact avec les deux amers.

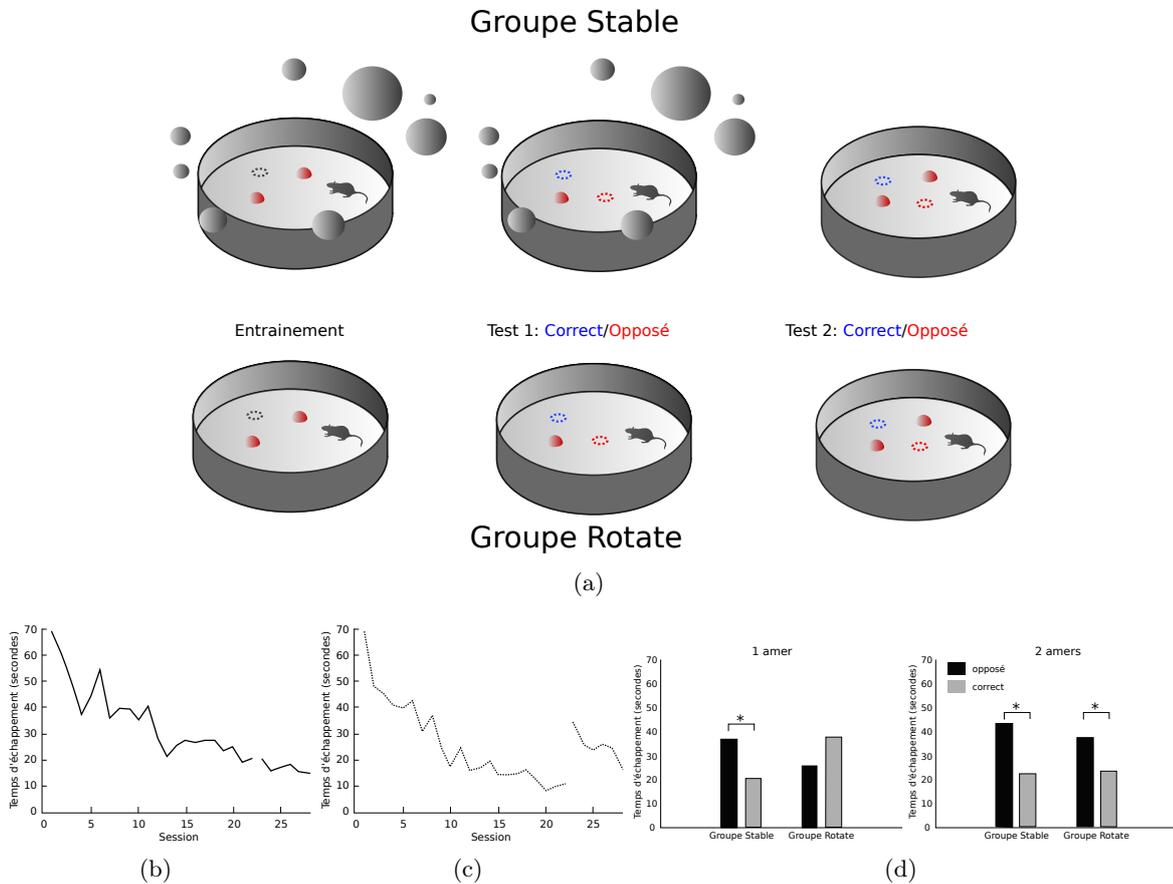


Figure 6.14 – **Protocole et résultats originaux de l’expérience de McGregor et al (2004a)**. (a) Protocole de l’expérience de McGregor et al (2004a). Voir texte pour détails. (b) Rappel de l’apprentissage moyen des groupes Rotate et (c) apprentissage des groupes Stable de l’expérience 1 de McGregor et al (2004a). L’expérience 2, que nous reproduisons ici, présente selon les auteurs les mêmes performances sur les 15 premières sessions. (d) Résultats observés aux essais de tests des groupes Stable et Rotate. Alors que le groupe Stable présente une différence significative de temps d’échappement dans les positions correcte et opposée de la plateforme à la fois pour 1 et 2 amers, le groupe Rotate ne présente ces différences que dans la condition 2 amers.

Rappel des résultats du groupe Rotate (section 5.1.3.2)

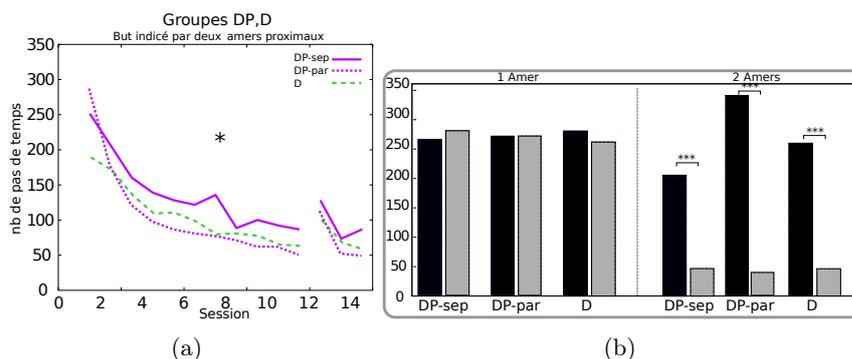


Figure 6.15 – **Simulation de l'expérience de McGregor et al (2004a)**. Rappel des résultats des groupes Rotate avec expert D et P (les groupes avec experts Taxon ne reproduisaient pas les résultats) (a) Courbes d'apprentissage; (b) Temps d'échappement pour les conditions 1 ou 2 amers (ils échouent dans la condition 1 amer).

6.3.2 Protocole et résultats simulés

L'environnement est reproduit avec les dimensions indiquées dans McGregor et al (2004a) : Le diamètre de la piscine est 2 m et celui de la plateforme 10 cm. Deux amers A et B, de 8 et 12 cm de diamètre, sont distants de 30 cm et la plateforme est située au sommet du triangle équilatéral composé avec les amers. Le nombre de sessions et d'essais, ainsi que le protocole sont identiques à ceux de l'expérience originale. Nous comparons les temps d'échappement des animaux dans les 4 essais de test.

Nous simulons ici les groupes constitués d'experts Direction et Planification (DP-sep, DP-par, D). Nous ajoutons néanmoins un expert Local (DL-sep, DL-par) et deux groupes L et P, qui montreront la nécessité ou l'inutilité d'une stratégie de guidage.

Nous avons déterminé précédemment que les groupes Rotate, ne pouvant utiliser les amers distaux (un rideau noir occultait la pièce), choisissaient une stratégie où un expert Direction (et non Taxon) ayant accès aux deux amers proximaux (en fait, deux experts Direction ancrés chacun sur un amer proximal) réussissait à apprendre la position d'une plateforme fixe en fonction de ces deux amers, et non d'un seul (Figure 6.15). La présence d'un expert Planification construisant sa carte avec odométrie ou avec les deux amers ne modifiait pas fondamentalement les caractéristiques des résultats. Nous avons conclu qu'une stratégie de guidage procédurale était donc suffisante pour expliquer l'absence d'occultation des deux types d'amers. Nous avons mentionné aussi que le groupe DP-par, au vu de sa courbe d'apprentissage pendant les sessions test, semblait autant rendre compte des résultats observés (Figure 6.15 a).

Nous abordons dans cette section la simulation de groupes ayant accès aussi aux amers distaux.

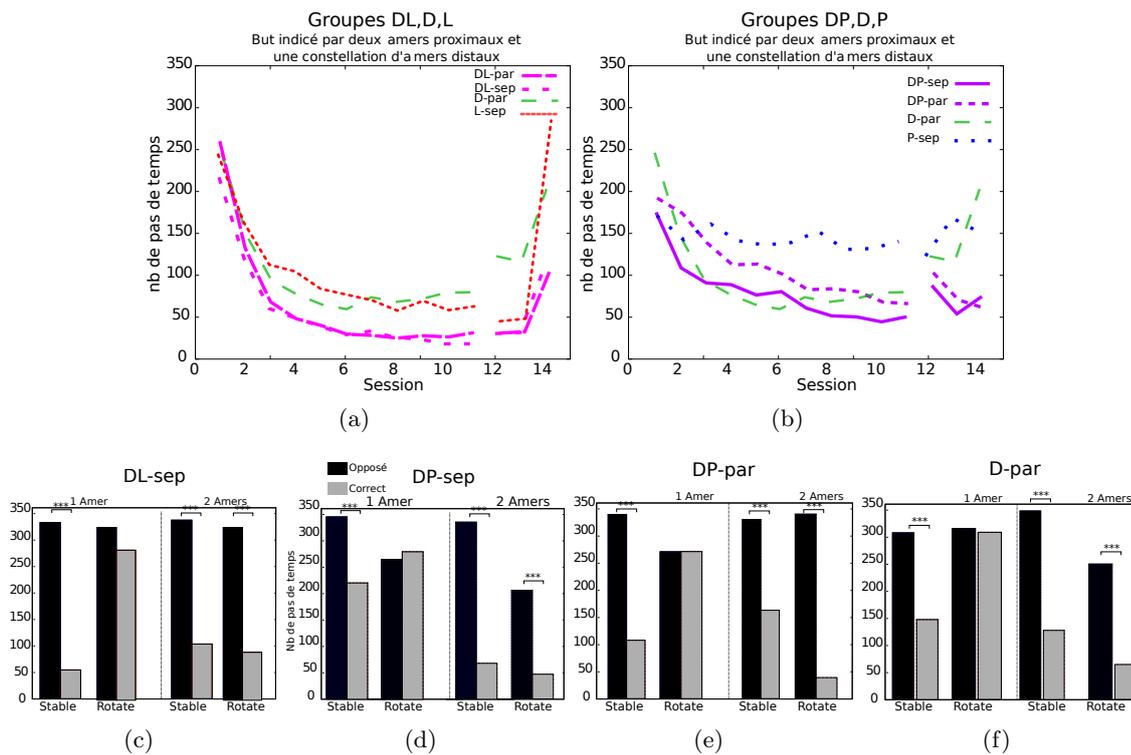


Figure 6.16 – **Simulation de l'expérience de McGregor et al (2004a).** (a,b) Courbes d'apprentissage des groupes Stable (a) DL-sep, D-par, D, L et (b) DP-sep, DP-par, D, P pendant l'apprentissage. (c-f) Temps d'échappement dans les positions correcte et opposée de la plateforme pour 1 et 2 amers, pour les groupes Rotate et Stable DL-sep, DP-sep, DP-par et D-par (les groupes Stable réussissent dans les deux conditions 1 et 2 amers).

6.3.2.1 Tous les groupes sont conformes aux comportements des animaux. Le groupe DP-par est le plus représentatif.

A l'instar des résultats obtenus avec le groupe Rotate, on constate encore que tous les groupes Stable avec experts DP ou D, de même qu'avec expert L ajouté ici, peuvent apprendre la tâche (Figure 6.16 a, b) et réussissent les essais test, c'est-à-dire qu'ils retrouvent à la fois la plateforme avec l'amer proximal A (et amers distaux) et les amers proximaux A et B (sans amers distaux) (Figure 6.16 c-f).

Les groupes D-par et DP-par, dans la condition Rotate, étaient les meilleurs candidats pour expliquer le comportement des animaux. Pour les groupes Stable, c'est le groupe DP-par qui s'en rapproche le plus, toujours au vu de l'amélioration des performances dans les sessions test – tous les autres groupes ayant de plus mauvaises performances à la session sans amers distaux (le groupe L-sep présentant la plus mauvaise performance pour retrouver la plateforme avec 2 amers proximaux, sans amers distaux). Nous allons donc commencer par analyser ce groupe

avant de commenter les autres.

6.3.2.2 Groupe DP-par. Apprentissage pendant l'entraînement et les sessions test

Le groupe parvient à de bonnes performances en fin d'entraînement. Le groupe DP-par apprend lentement, mais obtient les mêmes performances que le groupe D-par en fin d'entraînement. L'interaction avec l'expert Planification a donc simplement ralenti cet apprentissage. Il parvient au meilleur temps d'échappement de tous les groupes, lors de la session 14 sans amers distaux (Figure 6.16 b).

Tous les experts sont sélectionnés en fin d'entraînement. Les experts ont des taux de sélection qui évoluent pendant l'entraînement (Figure 6.17 c) et deviennent similaires en fin d'entraînement. Après avoir été sélectionné en début d'expérience, l'expert Planification a été dévalué au profit des experts Direction (amers proximaux et distaux), puis il a regagné de la valeur associative. On remarque donc que l'absence d'occultation n'intervient qu'au cours de l'expérience, vers la quatrième session, suite à l'apprentissage plus long des amers distaux par les experts de guidage, qui deviennent efficaces à ce moment-là. En fin d'entraînement, tous les experts sont donc jugés pertinents pour une sélection en présence ou non d'amers distaux.

Dans les sessions test, tous les experts sont recrutés. Le groupe DP-par Stable réussit les essais test avec une sélection conjointe des deux types d'experts, renforçant la sélection de l'expert de lieu pour la position opposée de la plateforme, et renforçant la sélection des experts de guidage pour la position correcte (Figure 6.17 e, premier et troisième groupes). La Figure 6.18 illustre un exemple de valeur de sélection acquise par les différents experts Direction au cours du test 2 amers, étayant le rôle des amers proximaux dans l'atteinte de la plateforme.

En considérant les sélections d'experts des groupes DP-par, Stable et Rotate (Figure 6.17 a, b, comparées à c, e), on constate que la présence d'amers distaux a modifié fondamentalement les sélections d'experts. La sélection conjointe des deux amers proximaux s'exerce à un taux beaucoup plus élevé dans le groupe Rotate, cela empêche ce groupe de réussir à trouver la plateforme dans les positions opposées ou avec un seul amer.

On constate aussi que la présence d'un expert Planification renforce la sélection conjointe entre experts Direction. En effet, dans le groupe D-par que nous allons analyser maintenant, les experts Direction ancrés sur les amers distaux deviennent beaucoup plus saillants que les autres et que dans le groupe DP-par.

6.3.2.3 Autres groupes

Dans le groupe D-par, les experts ancrés sur les amers distaux remplacent l'expert de lieu. La Figure 6.17 d montre effectivement que les experts Direction ancrés sur les amers distaux prennent de l'importance et contribuent à l'essai test 1 amer avec amers distaux (Figure 6.17 e, deuxième groupe). Ils dévaluent néanmoins les experts Direction A et B pendant les essais tests (Figure 6.17 d, sessions 12-14), mais ils ont été suffisamment entraînés pour être recrutés pendant l'essai test 2 amers (Figure 6.17 e, quatrième groupe). C'est la dévaluation de

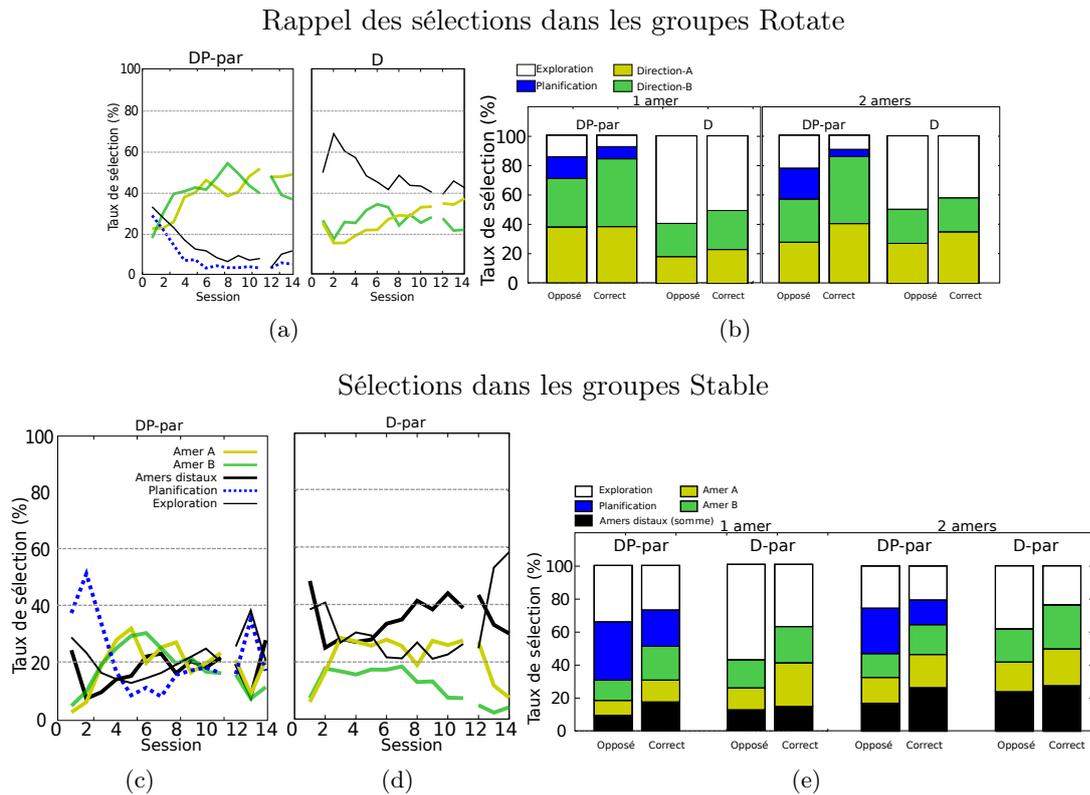


Figure 6.17 – **Simulation de l'expérience de McGregor et al (2004a)**. (a, b) Rappel des taux de sélection des experts pendant l'apprentissage dans les groupes Rotate DP-par et D et sélection des experts de ces groupes pendant les essais de test, conditions 1 ou 2 amers. (c-, d) Taux de sélection des experts pendant l'apprentissage dans les groupes Stable DP-par et D-par et (e) sélection des experts de ces groupes pendant les essais de test, conditions 1 ou 2 amers.

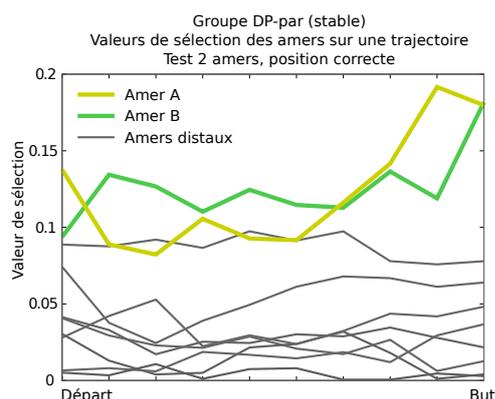


Figure 6.18 – **Simulation de l'expérience de McGregor et al (2004a)** Exemple des valeurs de sélection des amers proximaux et distaux des experts Direction dans le groupe DP-par, au cours d'une trajectoire allant vers la position correcte du but pendant l'essai test 2 amers.

ces amers proximaux qui a conduit ce groupe à une performance moins bonne que le groupe DP-par dans cet essai test. La présence d'amers distaux a donc contribué à une occultation des amers proximaux, toutefois pas suffisante pour modifier les résultats aux sessions test. Une prolongation de l'entraînement, renforçant encore l'ancrage sur les amers distaux, aurait peut-être abouti à un effet occultant.

Dans le groupe DP-sep, l'expert de lieu est à lui seul efficace dans l'essai 1 amer. Les experts Direction augmentent conjointement leur sélection dans le groupe DP-sep (Figure 6.20 a). Mais avec les amers distaux, l'expert Planification pourrait être à lui seul efficace pour positionner la plateforme avec un seul amer proximal. Son taux de sélection pendant cet essai est d'ailleurs plus important que dans la condition 2 amers (Figure 6.20 c, premier et troisième groupes). Pour vérifier cette hypothèse, nous avons simulé un groupe P-sep Stable, ayant accès aux seuls amers distaux. On constate en effet qu'après le même entraînement que le groupe DP Stable, ce groupe passe un temps significativement plus important dans le quadrant "correct" que dans le quadrant "opposé" (Figure 6.19).

La sélection de l'expert Planification dans le groupe DP-sep n'a pas empêché la sélection des experts Direction concernant les amers A et B (l'augmentation du taux de sélection de ces experts en fait preuve pendant l'entraînement), qui peuvent donc être recrutés lors de l'essai avec 2 amers, sans amers distaux (Figure 6.20 c, troisième groupe).

Dans les groupes DL-sep et DL-par, l'expert Local prend très majoritairement le contrôle. L'apprentissage s'effectue en environ 4 sessions et aboutit aux performances les meilleures pendant l'entraînement (Figure 6.16 a). En revanche, l'essai test avec 2 amers est peu réussi, d'ailleurs sans le concours de l'expert Local, à juger sur sa très mauvaise performance.

La Figure 6.20 b explique ce résultat, car on constate que le module de sélection accorde une confiance très importante à l'expert Local. L'essai avec 1 amer est réussi avec son concours

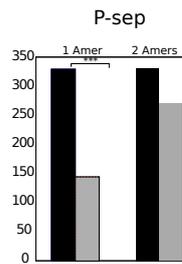


Figure 6.19 – **Simulation de l'expérience de McGregor et al (2004a)**. Temps d'échappement dans les positions correcte et opposée de la plateforme pour 1 et 2 amers, pour le groupe Stable P-sep.

(Figure 6.20 c, deuxième groupe), mais l'essai avec 2 amers a nécessité le recrutement des experts Direction (Figure 6.20 c, quatrième groupe), que le module de sélection a pu réaliser rapidement.

Ce groupe illustre le cas de sélection d'experts qui pourrait correspondre à une occultation d'amers proximaux par les amers distaux, mais l'apprentissage parallèle des experts ancrés sur les amers proximaux a suffi à ce qu'ils soient de nouveau candidats pour un recrutement.

6.3.2.4 Dans aucun groupe l'approche à un ou deux amers signalée par les auteurs n'a été retrouvée.

Malgré des analyses fines d'approche de la plateforme, nous n'avons pas pu apporter une explication quant à la différence de comportement des animaux reportée par les auteurs en présence et en l'absence d'amers distaux soit, respectivement, avoir tendance à ne toucher qu'un seul amer ou les deux, en se dirigeant vers la plateforme. Aucune autre analyse de sélection des amers proximaux n'a réussi à montrer cette tendance, quoiqu'on puisse imaginer que l'utilisation d'une carte cognitive nécessite moins l'approche des deux amers que deux experts Direction ancrés chacun sur l'un deux.

6.3.3 Discussion partielle

A l'instar des groupes Rotate, on retrouve l'absence d'occultation observée par les auteurs dans tous les groupes Stable mentionnés ici (avec experts D, L ou P). Le groupe DP-par semble correspondre plus précisément aux comportements des animaux, car il reproduit mieux les courbes d'apprentissage pendant la Phase 2. Cependant des groupes constitués d'experts directionnels procéduraux ont des résultats aussi très satisfaisants.

On remarque que tous ces groupes Stable parviennent à cette même absence d'occultation malgré des sélections d'experts très diverses, mais leurs apprentissages parallèles les ont rendu efficaces dans les conditions adéquates.

Dans leur discussion, les auteurs tentent d'expliquer cette absence d'occultation d'amers proximaux dans les deux groupes – résultats contraires à d'autres observations – par différentes hypothèses.

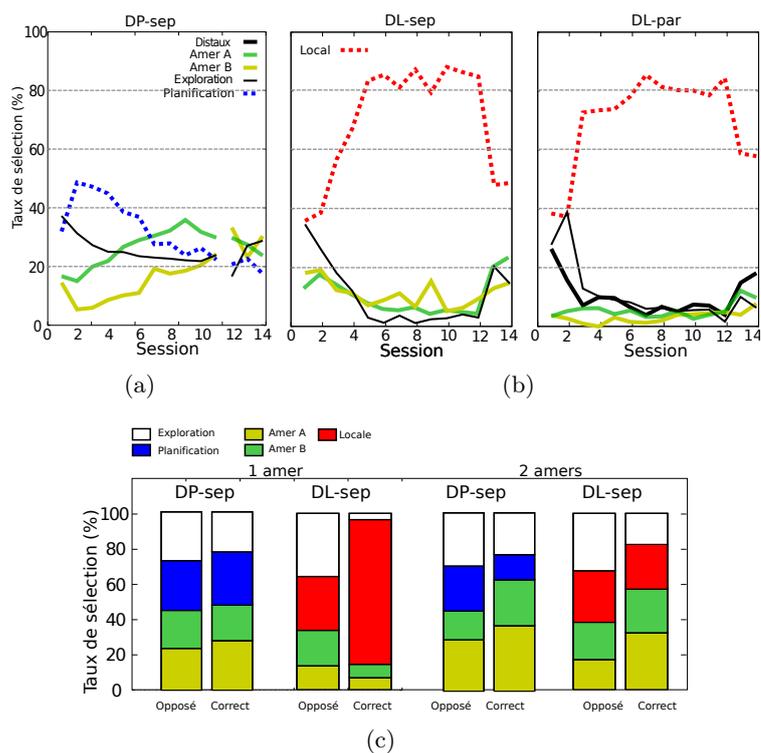


Figure 6.20 – **Simulation de l'expérience de McGregor et al (2004a)**. (a, b) Taux de sélection des experts pendant l'apprentissage dans les groupes Stable (DP-sep et DL-sep,-par). (c) Sélection des experts pour ces groupes pendant les essais de test, conditions 1 ou 2 amers.

Par exemple, ils évoquent la possibilité que les animaux fixent leur attention tantôt sur certains amers, tantôt sur d'autres, ce qui ferait progresser l'apprentissage de chaque amer indépendamment des autres. C'est ce que nous proposons : les amers proximaux et distaux peuvent être appris en parallèle par les experts de lieu et aussi de guidage dans les groupes DP-par et D. Les experts de guidage choisissent les amers séquentiellement, selon leur pertinence. Par rapport aux résultats obtenus dans les simulations de l'expérience de Roberts et Pearce (1999), où l'amer situé sur le but proximal avait tout de suite pris de l'importance, on peut avancer que la présence de deux amers proximaux, situés à côté du but, inciterait les animaux à apprendre à le positionner non seulement avec ces amers, mais aussi avec des amers plus lointains.

Les auteurs ont émis aussi l'hypothèse que le triangle formé par les deux amers proximaux et le but aurait mis en œuvre un module géométrique qui, selon certains auteurs, serait impénétrable à des effets associatifs de blocage ou d'occultation. Nous écartons cette explication, car les simulations montrent que l'absence d'occultation peut se retrouver dans un groupe composé uniquement d'experts de guidage – donc sans module géométrique a priori déterminé ou bien intégré dans une carte cognitive.

Enfin, les auteurs convergent vers l'utilisation d'une stratégie "heading vectors", différente d'une stratégie qui serait efficace uniquement avec un amer situé sur le but, ce que nous avons étayé dans la première approche de cette expérience. Cependant nous avons suggéré que la présence d'un expert Planification pouvait renforcer la sélection conjointe de tous les experts (groupes DP-par comparé à D-par) – et donc l'attention à tous les amers, A et B compris.

Comparées aux groupes Session-Same de Roberts et Pearce (1999) qui apprenaient à positionner dans des conditions similaires une plateforme balisée (Figure 6.10), les sélections d'experts des groupes Stable sont radicalement différentes : chez les Session-Same, un expert de lieu est sélectionné en début d'expérience, mais la balise prend très vite une importance supplantant les autres amers (Figure 6.10 b gauche haut). Chez les Stable, la compétition entre experts semble plus grande, car tous continuent à acquérir de l'importance et les amers distaux ne sont pas négligés (Figure 6.17 c). L'absence d'occultation serait ainsi provoquée par une "indécision" des animaux à choisir quels amers seraient les plus fiables, ce qui permet de conserver leur utilité dans des circonstances où ils peuvent servir à positionner le but.

Nous ne pouvons cependant proposer une utilisation séparée ou partagée des amers car, à part une amélioration des performances en sessions test observée chez les animaux et exprimée aussi par le groupe DP-par, rien dans les résultats des essais test eux-même ne suggère d'opter pour l'une ou l'autre des conditions. Proposons donc, comme au terme des expériences concernant le groupe Rotate, une hypothèse de parcimonie : l'absence d'occultation des amers proximaux par les amers distaux pourrait être expliqué dans cette expérience par le seul concours d'experts de guidage soumis à une règle associative. Mais cela veut dire *qu'une absence de blocage peut donc émerger aussi d'une telle interaction.*

6.4 Conclusion du chapitre

Nous rappelons que trois expériences faisant référence à certains effets de l'apprentissage spatial résultant de la loi associative, – deux les confirmant (effets de gradient ; effet de blocage), une autre les infirmant (absence d'occultation)– ont été correctement reproduites par le modèle. Elles ont permis de retrouver les effets suivants (Tableau 6.1) :

Présence d'amers proximaux et distaux

- un amer proximal situé sur ou près du but est toujours plus saillant que les amers distaux. Il est d'autant plus saillant qu'il est près du but ;
- un amer proximal peut bloquer l'utilisation des amers distaux lorsqu'il est appris préalablement plus intensément que les autres et qu'il peut prédire la position du but de façon fiable ;
- deux amers proximaux ne sont pas occultés par les amers distaux lorsqu'ils sont appris simultanément autant que les autres et qu'aucun ne prédit la position du but plus que l'autre.

Nous ajoutons les précisions suivantes :

- Les effets de gradient associatif, dus à la distance entre un amer proximal et distal (Rodrigo et al, 2006), ou de blocage d'amers distaux par un amer proximal (Roberts et Pearce, 1999), seraient mieux expliqués par la sélection conjointe d'un expert de guidage procédural et d'un expert de lieu déclaratif, ancrés chacun sur des catégories d'amers différentes :
- Dans les groupes où les effets précédents se constatent, le premier trajet vers la zone de la plateforme est contrôlé par un amer proximal.
- Ensuite, si la plateforme est absente, les experts ancrés sur les amers distaux peuvent être à nouveau recrutés. Ces experts exerceraient des effets de compétition ou de coopération avec l'expert précédent. Contrairement aux conclusions des expériences originales, l'effet de blocage des amers distaux ne serait donc pas total et l'effet de gradient pourrait même en dépendre.

Les effets précédents ne peuvent pas se retrouver tels qu'ils ont été établis par une compétition d'amers soumis à une règle associative.

- En revanche, l'absence d'occultation de deux amers proximaux situés à côté du but par des amers distaux McGregor et al (2004a) ne nécessite *pas forcément une intervention d'un expert de lieu, ni d'une configuration de type géométrique entre amers et but*, un expert directionnel étant capable de trouver le but en s'ancrant séquentiellement sur les amers proximaux ou distaux.

Effets associatifs et cognitifs Bien qu'il soit périlleux de généraliser à partir d'expériences peu nombreuses et qui ont des contextes environnementaux différents, nous déduisons des points précédents qu'ils ne rejettent ni l'une ni l'autre des deux hypothèses associative et cognitive

concernant l'apprentissage spatial. Comme nous l'avons proposé en Introduction, chacun de ces effets est lié de façon importante au contexte environnemental et à l'historique des animats.

Il est indéniable que des compétitions entre amers émergent des interactions entre les experts (de guidage seulement, ou de guidage et de lieu) mis en présence. Cependant cela n'empêche pas ces experts de coopérer, lorsqu'ils deviennent compétents pour rejoindre un même but.

Dans le modèle, seul l'expert Planification intègre réellement les amers distaux ou tous les amers par une règle déclarative. Cependant il n'a pu prendre majoritairement le contrôle sur l'expert de guidage qu'en début d'expérience. En présence d'experts de guidage, il ne peut donc pas expliquer à lui seul les effets observés, mais il y participe.

La conception de l'expert de guidage correspond à une sélection séquentielle allocentrique d'amers proximaux et/ou distaux. Ces experts sont sélectionnés avant d'interagir avec l'expert de lieu et certains experts peuvent prendre de l'importance pendant un entraînement, suite à une compétition entre eux. Cependant, contrairement aux hypothèses associatives, ces experts peuvent également coopérer pour positionner la plateforme lorsqu'elle est plus difficile à repérer. C'est pourquoi une présence de blocage et une absence d'occultation peuvent émerger par les différences de contextes environnementaux.

L'hypothèse d'amers spécifiques traités par l'un ou l'autre des experts (prise en compte des amers proximaux par un expert de guidage et des amers distaux par un expert de lieu) peut être étayée car elle a pu rendre compte de ces trois expériences. Cependant la prise en compte de tous les amers peut aussi se révéler compatible avec de telles reproductions – du moins en la présence de nombreux amers distaux. Précisons encore une fois que cela ne veut pas dire que ces amers soient obligatoirement intégrés dans une carte cognitive, ils peuvent être utilisés soit séquentiellement soit globalement.

En résumé, le modèle permet l'apprentissage parallèle de la pertinence de différents amers, ce qui laisse la possibilité que les actions correspondantes soient recrutées si besoin est, à condition qu'elles soient suffisamment apprises. Ces amers vont acquérir une valeur relative à la pertinence du but par un apprentissage de type associatif (dans les modules de sélection entre experts de guidage ou entre experts de guidage et de lieu), qui leur confère des importances différentes, *mais pas forcément exclusives*. Un apprentissage fondé sur l'intégration de ces amers (experts de lieu) serait néanmoins indispensable pour tenir compte de situations où il faut réagir rapidement au déplacement ou à l'absence d'un but.

Par la simulation des deux expériences du chapitre suivant, nous nous proposons d'investiguer plus précisément ces dernières conditions. En outre, dans ces expériences, les auteurs émettent explicitement l'hypothèse de l'utilisation d'une stratégie de guidage et d'une stratégie de planification gérées par deux systèmes neuraux différents, respectivement non dépendant et dépendant de l'hippocampe, dont le deuxième est censé construire une carte cognitive. Contrairement aux

expériences précédentes, les deux systèmes sont supposés focaliser respectivement sur les amers proximaux et distaux, qui seront mis en situation de conflit.

Tableau 6.1 – Résumé des résultats simulés des groupes Contrôles de ce chapitre, selon les configurations d’amers et si le but est déplacé ou non pendant l’entraînement (Ent.) et les test (Test). Les expériences sont indiquées par le nom du premier auteur de l’article correspondant. Les groupes "Conforme" sont ceux qui reproduisent au plus près le comportement animal ; les groupes "Possible" sont ceux qui présentent quelques différences avec le comportement observé ; les "Non conforme" ne reproduisent pas les résultats. Les groupes "Non simulé" n’ont pas participé à ces expériences. Enfin les sous-catégories "Isol.", "Sep." et "Par." signifient si les simulations ont été réalisées avec, respectivement, un experts seul, des experts avec entrées séparées ou partagées.

Exp.	Amers	But déplacé		Conforme			Non conforme		
		Ent.	Test	Isol.	Sep.	Par.	Isol.	Sep.	Par.
Rodrigo	1 balise 1 dist	non	oui		DP		T,D	TP	TD,DP
Roberts	1 balise 8 dist	oui	non		DP		T,D	TP,DL	TP
McGregor (Stable)	2 prox 8 dist	non	oui			DP	T, P	TP	TP

Exp.	Amers	But déplacé		Possible			Non simulé		
		Ent.	Test	Isol.	Sep.	Par.	Isol.	Sep.	Par.
Rodrigo	1 balise 1 dist	non	oui				L	TL,DL	TL,DL
Roberts	1 balise 8 dist	oui	non			DP	L	TL	TL
McGregor (Stable)	2 prox 8 dist	non	oui	D	DP,DL		L	TL,DL	TL,DL

Chapitre 7

Interactions entre systèmes de guidage et de lieu en situations de conflit

Les deux expériences reproduites dans ce chapitre sont mentionnées par les auteurs comme des situations de compétition entre stratégies de guidage et de lieu, dont les apprentissages sont respectivement indépendants et dépendants de structures nerveuses liées à l'hippocampe. Les situations de compétition sont les suivantes :

- Compétition ponctuelle (Devan et White, 1999), en présence d'un but visible ou invisible et de plusieurs amers distaux ;
- Compétition graduelle (Pearce et al, 1998), en présence d'un but caché indicé et de plusieurs amers distaux.

Certains aspects de ces expériences prennent en compte des groupes de rats intacts (Contrôles) mais aussi des groupes de rats lésés de l'hippocampe (rats Hc, supposés ne produire qu'une stratégie de guidage) et du striatum dorsolatéral (rats DLS, supposés ne produire qu'une stratégie de lieu). Ces derniers permettront d'étayer la pertinence de nos experts, en comparant leurs résultats à ceux d'animats constitués d'une seule sorte d'expert, respectivement supposés équivalents aux groupes T, D, pour les rats Hc, et aux groupes L, P, pour les rats DLS.

Ces expériences ont déjà fait l'objet d'analyses détaillées dans Dollé et al (2008), Dollé et al (2010a) et Dollé et al (2010b). Nous reprenons ici les points essentiels et en développons d'autres.

7.1 Compétition ponctuelle (Devan et White, 1999)

L'hypothèse initiale de cet article était spécifiquement de démontrer les effets de lésions du fornix/fimbria (région hippocampique) et de différentes parties du striatum (DLS et DMS) dans une tâche où pouvaient être simultanément apprises des stratégies de guidage et de lieu (but visible), respectivement non-dépendantes et dépendantes de l'hippocampe. Outre cette investigation particulière, une stratégie de lieu devait s'exprimer lors de sessions particulières (but caché) chez les groupes Contrôle et DLS, et un test final (but visible déplacé) devait mettre les deux stratégies en compétition de façon à ce que la stratégie de guidage prenne le contrôle. Ce sont ces deux situations que nous essayons de reproduire ici, afin d'analyser les conditions d'interactions

des experts lors d'utilisations alternatives des deux types de stratégies.

7.1.1 Protocole et résultats observés

Dans l'expérience 1 de Devan et White (1999), la piscine a un diamètre de 172 cm et est disposée dans une pièce avec amers distaux. Trois groupes de rats – intacts (groupe Contrôle), lésés du fornix/fimbria (groupe Hc) et du striatum dorso-latéral (groupe DLS) – sont entraînés à mémoriser l'emplacement d'une plateforme dans une piscine durant neuf jours, à raison de 4 essais par jour. Les jours 1-2, 4-5 et 7-8, la plateforme est visible et les jours 3, 6 et 9, elle est cachée. Le dixième jour, il y a un seul essai de compétition, au cours duquel la plateforme visible est déplacée (Figure 7.1 a). Quatre principaux résultats concernent la problématique des interactions entre stratégies de lieu et de guidage, selon les lésions effectuées (contrôles, fornix/fimbria, DLS et DMS). Nous ne donnons que les résultats de groupes de rats que nous supposons pouvoir simuler et laissons de côté les rats lésés du DMS (Figure 7.1 b), dont les résultats sont similaires à ceux des animaux lésés du fornix/fimbria. En effet, les caractéristiques de ce groupe ne sont pas aisément modélisables, dans l'état actuel des connaissances biologiques et de nos possibilités de modélisation.

Nous introduisons ici les trois principales caractéristiques des résultats observés :

(i) Les rats Contrôle, Hc et DLS obtiennent les mêmes performances pour atteindre la plateforme visible durant l'entraînement, ce qui suggère aux auteurs que les stratégies dépendantes ou indépendantes de l'hippocampe peuvent réussir la tâche.

(ii) Contrairement aux autres groupes, le groupe Hc ne réussit pas à atteindre le but lorsque la plateforme est cachée. La stratégie dépendante de l'hippocampe serait donc nécessaire pour localiser la plateforme cachée.

(iii) En revanche, le jour de compétition, le groupe Hc atteint significativement plus vite la plateforme que les autres. L'utilisation d'une stratégie de lieu par le groupe Contrôle aurait donc pu diminuer ses performances. Les auteurs interprètent ce résultat en terme de compétition entre stratégie de lieu et de guidage (la lésion d'un des systèmes entraînant l'amélioration de la performance), les deux étant suffisamment apprises pendant l'entraînement.

Pour corroborer cette hypothèse, les auteurs notent que, pendant le test, les rats du groupe Contrôle peuvent être divisés en deux groupes sur la base de leur comportement. Ils calculent le nombre de rats qui se dirigent du point de départ jusqu'à 5 cm autour de l'emplacement de l'ancienne plateforme (les "place-responders") et les rats qui n'entrent pas dans ce périmètre (les "cue-responders") avant de se diriger vers l'emplacement courant. Sur 10 rats, ils dénombrent 6 "place-responders" et 4 "cue-responders". Les premiers auraient donc contribué à diminuer les performances du groupe.

7.1.2 Protocole et résultats simulés

L'environnement comprend une piscine de 172 cm avec plateforme visible ou invisible et 8 amers distaux placés pseudo-aléatoirement. Les jours 1-2,4-5,7-8, l'indice visuel de 10 cm de diamètre, représentant la plateforme visible, est situé au centre du quadrant sud-ouest de l'environnement. Les jours 3,6,9, l'amer est absent mais la zone de récompense reste présente au même endroit. Le dixième jour, la plateforme et l'amer visuel sont déplacés au centre du

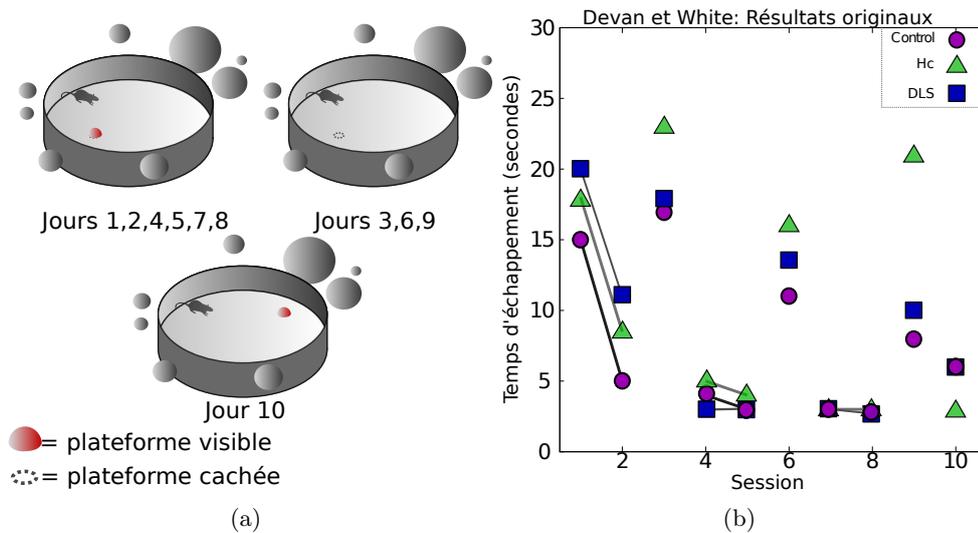


Figure 7.1 – **Expérience de Devan et al. (1999)**. (a) Protocole original (b) Résultats originaux. Les symboles joints représentent les sessions où la plateforme est visible.

quadrant nord-est de l'environnement. Au début de chaque essai, l'animat est placé sur le bord de la piscine, à l'un des quatre points cardinaux. Le jour de compétition, l'animat est placé à équidistance de l'ancienne et de la nouvelle position de la plateforme (Figure 7.1 a).

La plateforme étant quelquefois visible mais toujours fixe, des groupes simulés avec un expert Taxon et un expert Local seront aussi intégrés, en plus des experts Direction et Planification. Le groupe Contrôle sera représenté par les groupes TP, TL ou DP ; le groupe Hc par les groupes T ou D ; le groupe DLS par les groupes L ou P. Pour une cohérence des groupes, les groupes avec entrées séparées (-sep) (accès à la plateforme visible pour experts Taxon ou Direction, aux amers distaux pour experts Local ou Planification) seront testés ensemble, ainsi que les groupes avec entrées partagées (-par) (accès à la fois à la plateforme visible et aux amers distaux pour tous les experts).

7.1.2.1 Le groupe TL-sep est le plus conforme au comportement des animaux

Seuls trois groupes, avec expert Taxon (TL-sep et par et TP-sep), réussissent le test de compétition. Tous les groupes parviennent plus ou moins rapidement à rejoindre la plateforme pendant l'entraînement, en revanche, seuls les groupes avec Taxon/Local (sep et par) et un seul avec Taxon/Planification (sep) (Figure 7.2 b, c, d) ont des performances au test de compétition conformes à ceux des animaux. Le jour 10 en effet, le groupe T atteint significativement plus vite que les autres groupes la plateforme visible et déplacée (Figure 7.2 b, c, d). Les groupes Contrôle TL et TP ont des résultats intermédiaires, meilleurs que les groupes L ou P.

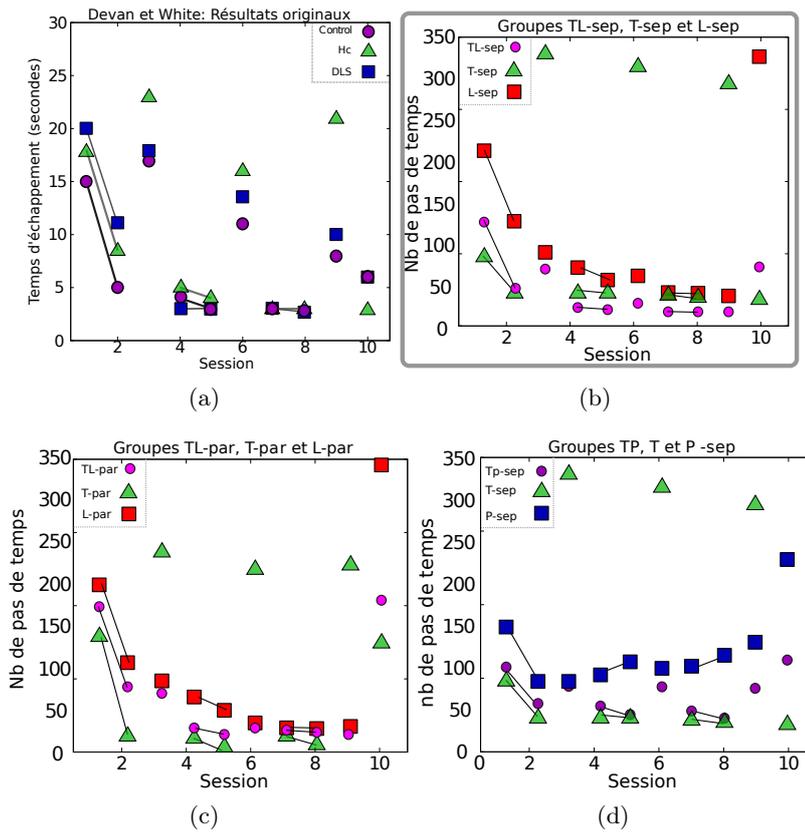


Figure 7.2 – Simulation de l'expérience de Devan et al. (1999). Résultats obtenus par les seuls groupes dont les résultats à l'essai de compétition sont conformes à celui des animaux : (a) Résultats originaux (b) groupes TL,T,L-sep;(c) groupes TL,T,L-par ; (d) groupes TP,T,P-sep .

Nous avons mis les Figures relatives à tous les autres groupes en annexe – sans représenter les résultats avec experts Direction et Local – qui ont le même profil (Figure E.1 b, c).

7.1.2.2 Comparaison des groupes où l'expert Taxon est avec l'un des experts de lieu, Local ou Planification.

Estimant qu'un apprentissage rapide de l'expert de lieu pouvait être nécessaire pour qu'il puisse parvenir à rejoindre la plateforme cachée le troisième jour, l'hypothèse de l'intervention d'un expert Local ne semblait pas adéquate. Les groupes avec Planification répondent en effet à tous les critères comportementaux des animaux, cependant ils ne parviennent pas à reproduire, les jours où la plateforme est invisible, l'amélioration des performances entre ces trois sessions observée chez les animaux (Figure 7.2 a).

Des simulations avec expert Local – dont, on le rappelle, les mécanismes d'apprentissage sont identiques à ceux de l'expert Taxon et rend l'acquisition plus lente que celle de l'expert Planification – ont pu représenter l'amélioration d'apprentissage précédemment citée.

Nos analyses vont concerner tout d'abord les comparaisons des groupes TL-sep et TP-sep (Figure 7.2 b, d), pour estimer l'impact des interactions de l'expert T avec l'un ou l'autre des experts de lieu.

7.1.2.3 Groupes TL-sep et TP-sep et groupes "lésés" associés.

Les jours où la plateforme est visible (1-2, 4-5, 7-8). Conformément aux résultats observés (groupes Contrôle et Hc), les groupes TL ou TP ont des performances non significativement différentes de celles des groupes T.

Les compétences similaires de ces groupes suggèrent que l'expert Taxon prend majoritairement le contrôle ces jours-là. Les taux de sélection de l'expert Taxon augmentent en effet pendant ces sessions (Figure 7.3 a, b), plus dans le groupe avec expert Local (Figure 7.3 f comparé à g, groupe "visible").

La plateforme étant visible, on aurait du s'attendre à une sélection beaucoup plus importante de l'expert Taxon mais, par la présence de trois sessions avec plateforme cachée, il ne peut pas prendre autant le contrôle sur l'expert de lieu ou l'exploration. La Figure 7.3 d montre en effet le groupe T-sep, dont le taux de sélection de l'expert Taxon n'augmente pas suivant les sessions, mais reste au même niveau tout en sélectionnant beaucoup d'exploration. L'absence de plateforme pendant trois essais contribue donc à maintenir un niveau de sélection d'exploration élevé. Comme les résultats des groupes TL et TP sont similaires à celles du groupe T, on peut en déduire que les deux experts de lieu jouent un rôle exploratoire.

Comme prévu, l'expert Local a un apprentissage plus lent, mais plus performant, que l'expert Planification. Ce dernier atteint environ 80 pas de temps en 2 sessions, contre 4 sessions pour l'expert Local. Cependant l'expert Planification reste sur sa performance, alors que l'expert Local peut l'améliorer au cours des sessions, en atteignant des niveaux semblables à ceux de l'expert Taxon – comme observé chez les animaux. Les deux experts de lieu ont des profils de

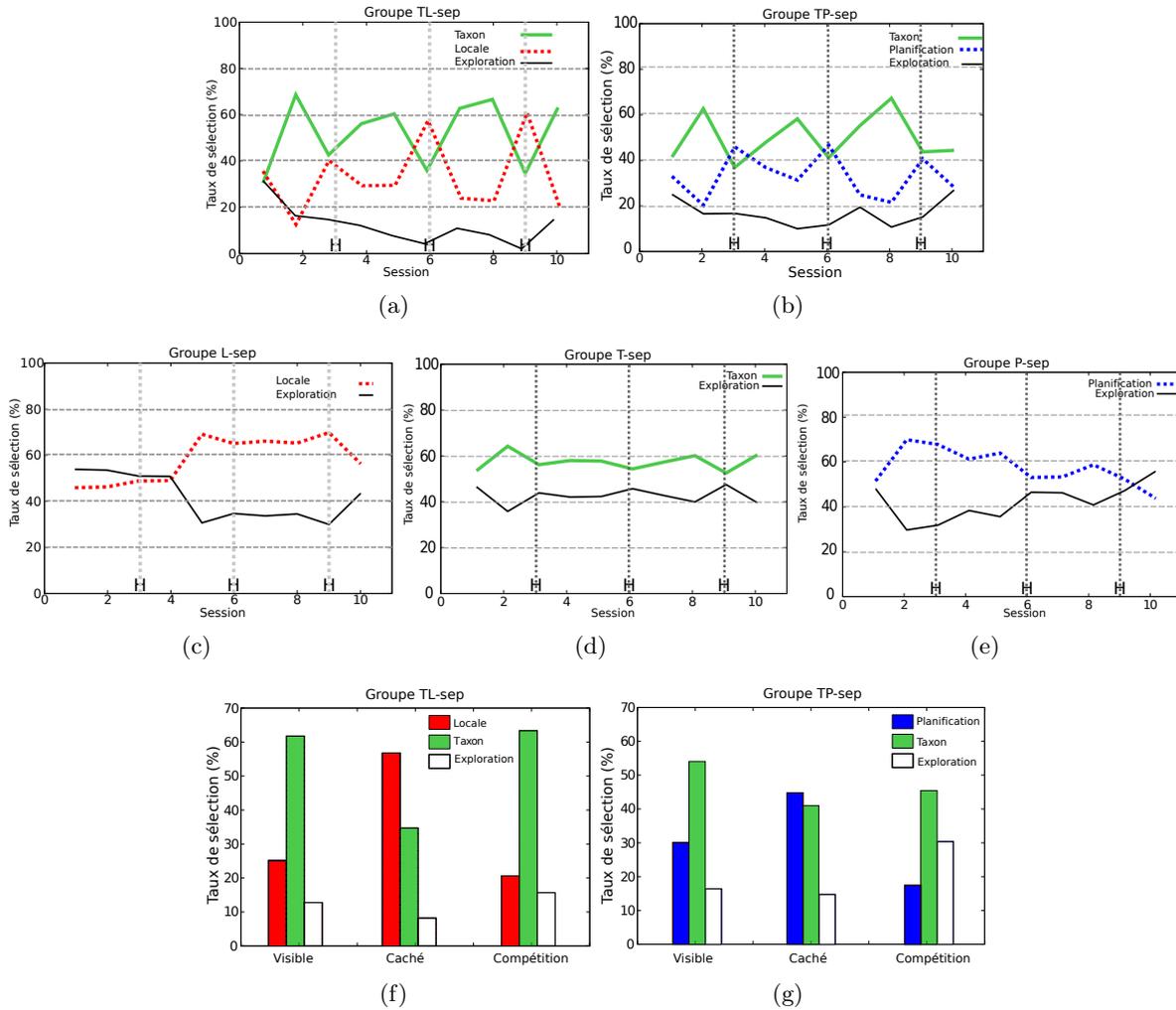


Figure 7.3 – **Simulation de l'expérience de Devan et White (1999)** (a-b) Taux de sélection des experts par session dans les groupes TL-sep. (c-e) Taux de sélection des experts par session dans les groupes L-sep, T-sep et P-sep.(f-g) Taux de sélection cumulés des experts TL-sep et TP-sep selon la configuration de l'environnement (but visible, caché et test de compétition).

sélection semblables dans les groupes Contrôle (Figure 7.3 a, b), mais pas dans les groupes où ils sont sans expert de guidage (Figure 7.3 c, e). L'apprentissage du Local lui permet, dès la quatrième session, d'avoir plus de chance d'être sélectionné que l'expert Planification.

Les jours où la plateforme est cachée (3, 6, 9). Conformément aux résultats observés (groupe Hc), le groupe T-sep ne réussit pas la tâche. Les animats utilisant un expert Taxon, comme les animaux supposés ne pas pouvoir construire de représentation spatiale, ne peuvent pas positionner la plateforme cachée bien avant le temps maximal, ce qui était prévisible (Figure 7.2 a, c). Il va donc se comporter dans les groupes Contrôle comme un expert exploration.

Comme les groupes d'animaux Contrôle et DLS, seuls les groupes TL, TP et L, P réussissent à retrouver la plateforme. Les compétences similaires des groupes TL/L et TP/P (Figure 7.2 b, d) suggèrent que l'expert de lieu prend majoritairement le contrôle ces jours-là dans les groupes Contrôle. La Figure 7.3 a, b indique en effet une augmentation du taux de sélection de ces experts, plus important pour l'expert Local (Figure 7.3 f, comparé à g, groupe milieu). C'est ce qui rend ses performances de plus en plus efficaces semblables à celles des animaux observés.

7.1.2.4 L'essai de compétition (jour 10).

Conformément aux résultats observés, le groupe T-sep (Hc) a les meilleures performances. Rappelons que dans ce test la plateforme est visible, mais déplacée. Malgré un apport important d'exploration, il est resté efficace pour diriger les animats vers le nouvel emplacement de la plateforme (Figure 7.3 d).

Les groupes L et P (DLS) mettent le temps le plus long pour rejoindre la plateforme. Dans la simulation en effet ces groupes ont les performances les moins bonnes pour rejoindre la plateforme. L'expert P n'a pas pu effectivement repositionner le nouvel emplacement de la plateforme pendant ce premier et seul essai. L'expert L a plus de difficulté, car il doit faire un réapprentissage complet de ce nouveau lieu (Figure 7.2 b, c, d).

Les groupes TL et TP présentent des performances intermédiaires. Ils mettent en effet significativement plus de temps que le groupe T, mais significativement moins que les groupes L et P, à rejoindre la plateforme. La Figure 7.3 a, b et f, g (groupes droits) montre que ces groupes sélectionnent l'expert Taxon, mais aussi l'expert de lieu. La présence d'un expert Local, par rapport à un expert Planification, renforce considérablement la sélection du Taxon (Figure 7.3 a comparée à b, et f comparé à g, droite). Des cartes de navigation des experts Taxon, Locale et Planification dans ces groupes, à la fin des sessions d'entraînement, illustrent comment ils ont appris à positionner très précisément la plateforme (Figure 7.4 a, b, c).

Mais, alors que l'expert Taxon est capable de diriger les animats vers le nouvel emplacement, les experts Local et Planification vont avoir tendance à les entraîner vers l'ancien. La chute de performance des groupes TL et TP serait donc due à ce "détournement" de trajectoire, que les auteurs avaient attribué à des animaux "place-responders" (6 animaux sur 10) qui se dirigeaient plus souvent que des "cue-responders" (4 animaux) vers la zone précédemment occupée.

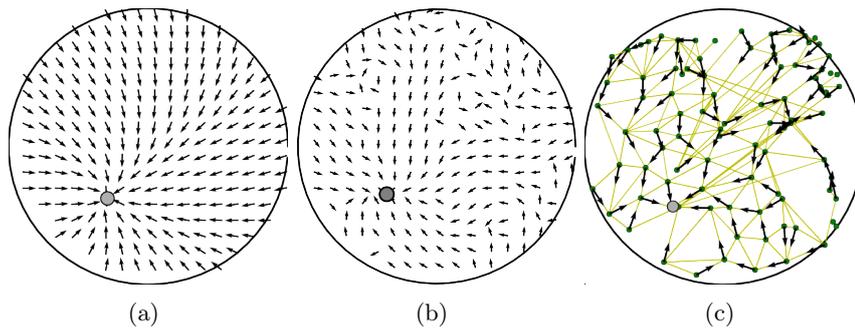


Figure 7.4 – **Simulation de l'expérience de Devan et White (1999)** Cartes navigationnelles des experts (a) Taxon, (b) Local et (c) Planification en fin d'entraînement (jours 8 et 9) dans les groupes TL-sep et TP-sep

Nous avons calculé avec le même critère les proportions de tels groupes parmi les animats. Dans le groupe TL (resp. TP), les "place-responders" représentent 40% (resp. 58 %) et les "cue-responders" 60% (resp. 42 %) de la population, ce qui est traduit par une plus grande sélection de l'expert Taxon au test pour le groupe TL et sa meilleure performance par rapport au groupe TP.

Les biais par une stratégie de lieu et par une stratégie visuelle indiqués par les auteurs pour ces animaux se retrouvent dans les taux de sélection respectifs des experts chez les animats place- et cue- responders (Figure 7.5 a, b). Les premiers sélectionnent, comme prévu, plus significativement l'expert de lieu, les seconds plus significativement l'expert de guidage et ce, quel que soit l'expert de lieu. Des trajectoires typiques de chacun des deux groupes sont illustrées en Figure 7.5 c, d.

Des analyses plus détaillées de l'historique de ces deux groupes d'animats – performances respectives des experts au cours des sessions – n'ont pas pu encore élucider le déterminisme de ces différences.

7.1.2.5 Autres groupes (TL-par, TP-par, DP-sep, DP-par).

Il faut distinguer ici les groupes associés à TL-par des trois autres groupes. En effet, seul les premiers ont un comportement conforme à celui des animaux au test de compétition.

L'expert Local du groupe TL-par prend exclusivement le contrôle pendant l'entraînement. Ayant accès aux amers distaux et à la plateforme visible, il peut donc conduire les animats vers la plateforme avec le seul expert Local (cf. le taux de sélection très faible de l'expert Taxon), quelle que soit la condition environnementale (Figure 7.6 a). Il a cependant plus de difficulté au test de la plateforme déplacée, ce qui fait que le groupe TL-sep reste le groupe ayant les résultats les plus proches de ceux des animaux.

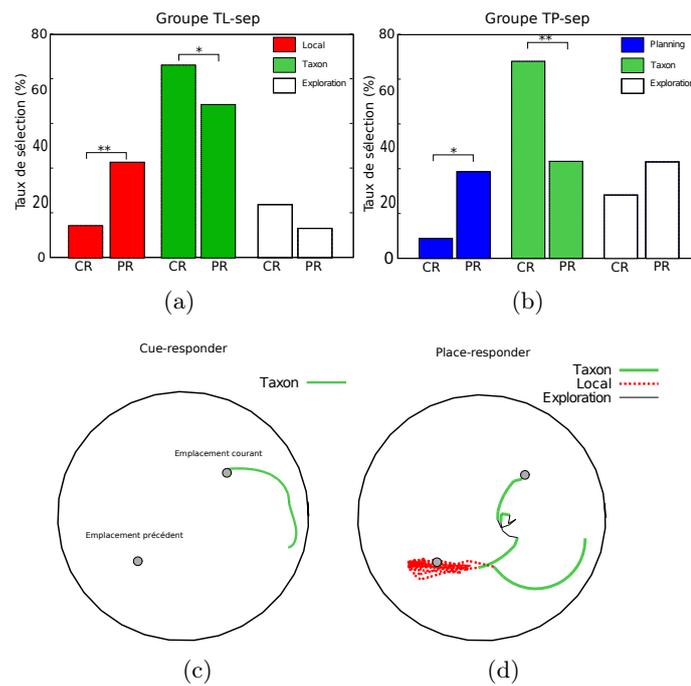


Figure 7.5 – **Simulation de l'expérience de Devan et White (1999)** (a-b) Taux de sélection des experts des groupes TL-sep et TP-sep dans le test de compétition selon les individus “cue-responders” et “place-responders”. (c-d) Trajectoires typiques d'un individu “cue-responder” (CR) et d'un individu “place-responder” (PR).

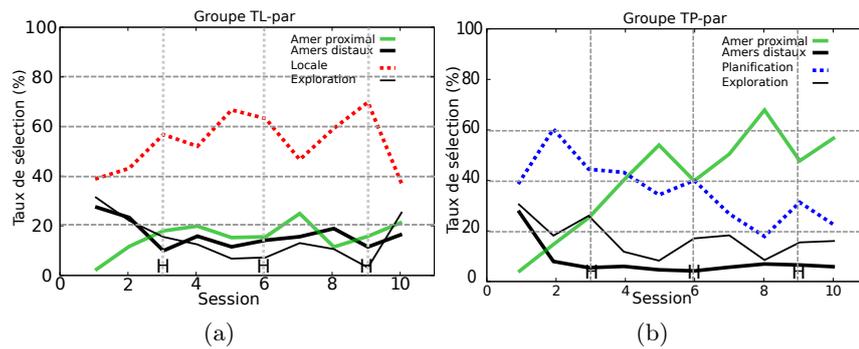


Figure 7.6 – **Simulation de l'expérience de Devan et White (1999)**. Taux de sélection des experts par session dans les groupes TL-par et TP-par.

Le Taxon devient beaucoup trop compétent. La Figure 7.6 a et b illustre les interactions différentes de l'expert Taxon avec les deux experts de lieu : dans le groupe avec expert Local, l'expert Taxon ne parvient pas à prendre le contrôle, ce qui est l'inverse lorsqu'il se trouve avec un expert Planification – et cela cause sa trop bonne performance au test de compétition. Les experts Taxon et Local ayant tous deux accès à tous les amers et le même type d'apprentissage, le module de sélection favorise donc l'utilisation d'une intégration d'amers versus une utilisation non intégrée.

Dans les groupes DP-sep et DP-par, l'expert Direction est beaucoup trop dévalué (cf Annexe E pour les Figures). Les groupes avec expert Direction n'ont pas réussi à reproduire l'essai de compétition. Malgré une meilleure sélection que celle du Taxon lorsque l'expert Direction est seul (Figure E.2 e), il réagit plus difficilement au déplacement de la plateforme. Contrairement à l'expert Taxon, cet expert ne parvient jamais à avoir un meilleur taux de sélection que l'expert de lieu pendant l'entraînement (Figure E.2 c, d), ce qui rend impossible sa prise de contrôle au test de compétition. La Figure E.2 f des taux cumulés d'experts montre effectivement un fort taux de sélection de l'expert Planification pendant les trois conditions expérimentales.

7.1.3 Discussion partielle

Le changement entre stratégies de guidage et de lieu d'une session à l'autre, tel qu'il a été observé pendant les sessions d'entraînement par Devan et White (1999) lorsqu'une plateforme fixe dans une piscine de Morris est visible ou invisible, a pu être reproduit par des groupes constitués d'experts de guidage égocentré ou allocentré (Taxon ou Direction), et de lieu procédural ou déclaratif (Local ou Planification). En revanche, le comportement pendant le test de compétition des groupes d'animaux Contrôle et lésés a été le mieux reproduit par des animats utilisant un expert Taxon et un expert Local, tous deux procéduraux, et chacun ancrés sur leurs amers spécifiques. Cela corrobore l'hypothèse de systèmes parallèles de navigation supposée par les auteurs de cette expérience, *mais non la nécessité d'utiliser une carte topologique*.

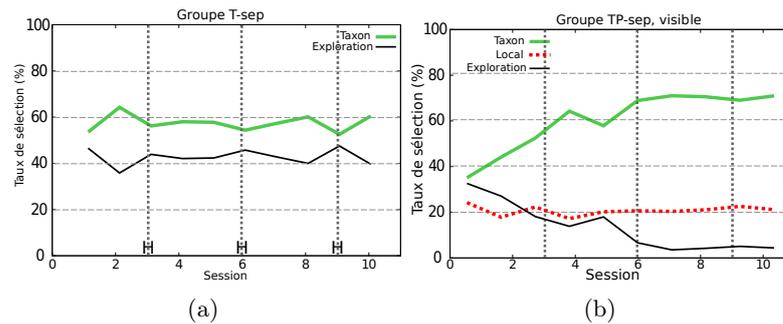


Figure 7.7 – **Simulation de l'expérience de Devan et White (1999)** Comparaison des sélections d'experts dans le groupe TL-sep (a) avec 3 sessions avec plateforme invisible et (b) sans ces 3 sessions (i.e., la plateforme est donc toujours visible).

Le conflit entre amers proximal et amers distaux n'a été résolu que par un groupe constitué d'une stratégie de guidage egocentrée et non allocentrée. Les interactions entre ce type d'expert et un expert de lieu aboutissent en effet à une meilleure saillance accordée à l'amer proximal en fin d'entraînement (au lieu d'une saillance plus grande accordée aux amers distaux avec un expert de guidage allocentré). Cependant, les amers distaux n'y ont pas été occultés, ce qui a permis le changement de stratégie lors des essais avec plateforme invisible.

L'interaction avec un expert Local s'est imposée car, contrairement à l'expert Planification, il a été capable d'améliorer progressivement ses performances les jours où seuls les amers distaux sont disponibles. L'adéquation de ce type d'expert de lieu dans cette condition expérimentale n'est pas surprenante puisque la plateforme est toujours restée à la même place pendant l'entraînement. En outre, cette interaction a instauré un niveau de compétition plus important entre experts de guidage et de lieu. Ce niveau est dû à la similitude de leurs mécanismes d'apprentissage, qui ont ainsi favorisés la reproduction du comportement animal au test de compétition.

Cette expérience a mis en valeur également les modifications importantes d'interactions entre stratégies dues à la seule présence des trois sessions pendant lesquelles la plateforme est cachée. Local et Taxon sont alternativement sélectionnés selon la visibilité du but, nous avons simulé le même nombre de sessions et le même test de compétition, mais en laissant la plateforme toujours visible. Les sélections d'experts sont considérablement modifiées. L'expert Taxon prend un contrôle de plus en plus important (Figure 7.7 b comparé à a). En conséquence, le groupe TL-sep (équivalent Contrôle) devient aussi performant que le groupe T-sep (équivalent Hc) pendant l'essai de compétition : en effet, il met un nombre de pas de temps égal à 40 (le groupe T-sep en ayant mis précédemment 30), alors qu'avec ces trois sessions avec plateforme invisible il en mettait 80 pour rejoindre la plateforme déplacée.

La stratégie de lieu a donc bien perturbé la trajectoire uniquement par son entraînement supplémentaire dans ces trois sessions. Cela pourrait être aisément confirmé par expérience. Cette

sensibilité des sélections d'experts au contexte avait déjà été signalée dans la section précédente.

Certains résultats observés n'ont pas pu être exactement reproduits par les simulations. En effet on note une bien plus mauvaise performance du groupe DLS (simulé par nos groupes L-sep ou P-sep) lorsque la plateforme est cachée (Figure 7.2 a comparée à 7.2 b-d).

Cela est dû au fait que les experts de lieu convergent trop rapidement vers leurs performances maximales, que la plateforme soit visible ou non. Le groupe DLS au contraire a un apprentissage moins performant, mais semble avoir gardé un sens lié à la vision de la plateforme, ce qui améliore ses performances lors des sessions avec plateforme visible. Aucun des deux experts Local ou Planification ayant accès à tous les amers – y compris l'amer proximal – n'est parvenu à reproduire ce comportement. On pourrait alors supposer que les animaux aient sélectionné une stratégie Locale pendant les tests à plateforme visible et une stratégie Planification dans les autres (les performances de la Planification étant inférieures après la troisième session), mais une tentative de mise en présence des trois experts n'a pas abouti à plus de conformité avec les résultats observés (les résultats sont similaires à ceux obtenus par le groupe TL-sep).

Malgré ces limitations, l'hypothèse des auteurs supposant une relation compétitive entre DLS et Hc – i.e., entre une stratégie de guidage et une stratégie de lieu, ayant construit une représentation spatiale –, a été corroborée par les simulations. Elle n'a pas été retrouvée pendant l'entraînement, ce à quoi nous nous attendions, puisque les amers distaux et l'amer proximal n'étaient pas en conflit. En revanche, la situation aurait pu aboutir à une coopération entre experts, mais elle ne s'est pas mise en place. On ne remarque en effet pas d'améliorations significatives inter-sessions ni d'amélioration de la performance des groupes (équivalents) Contrôle par rapport aux groupes (équivalents) lésés. Les deux stratégies n'étant pas vraiment complémentaires – puisque positionner une plateforme visible est facilement réalisable par l'une ou l'autre stratégie –, il est possible qu'une telle interaction de coopération n'ait pas émergé.

Les auteurs ne la signalent d'ailleurs pas, car ils supposent qu'une interaction entre DLS et Hc ne peut être que compétitive, mais qu'elle pourrait être coopérative entre les structures nerveuses DMS et Hc. Nous ne pouvons vérifier cette hypothèse, ayant délibérément omis de modéliser ce groupe.

En revanche, l'interaction compétitive entre Hc et DLS était supposée être traduite par les auteurs par une compétition entre stratégies de guidage procédurale et stratégie de lieu déclarative. Nous avons montré qu'une représentation intégrant les amers pouvait en effet être en compétition avec une stratégie visuelle égocentrée procédurale, mais que la représentation spatiale de la stratégie de lieu pouvait être aussi utilisée de façon procédurale comme l'expert Local. Une telle possibilité n'a pas été investiguée par les auteurs, alors qu'elle avait pu être supposée par Granon et Poucet (1995), impliquant l'hippocampe mais pas des zones du cortex médial préfrontal (nous reviendrons sur ce point en Discussion générale). Des expériences d'inactivation de l'hippocampe et/ou de zones correspondantes de ce cortex, donnant ou non des résultats semblables pour les deux groupes, pourraient clarifier ce point.

L'expérience suivante apportera des compléments d'information sur les interactions d'experts lors de conflits répétitifs entre un amer proximal et des amers distaux.

7.2 Compétition graduelle (Pearce et al, 1998)

L'hypothèse initiale de cet article était de démontrer que des rats dont l'hippocampe est lésé (groupe Hc) peuvent réussir autant que des rats intacts (groupe Contrôle) une tâche de navigation avec plateforme invisible indicée et déplacée à chaque session. Contrairement à la situation de l'expérience précédente, à chaque déplacement les stratégies vont se trouver en conflit, puisque ce déplacement modifie la position du but vis-à-vis des amers distaux, mais pas vis-à-vis de l'amer proximal, qui se trouve toujours à une même distance et orientation. Les auteurs font l'hypothèse que, chez les rats Contrôle, une stratégie de guidage et une stratégie de lieu vont se trouver en compétition tout au long de l'expérience, mais que la première va prendre progressivement le contrôle sur l'autre.

Les auteurs ajoutent à cette expérience un groupe de rats intacts, mais mis dans une situation où ils sont censés ne pas pouvoir construire une représentation spatiale avec les amers distaux (groupe Trial), faisant l'hypothèse que ce groupe va se comporter comme un groupe Hc.

La plateforme étant indicée, notre hypothèse est qu'un expert Direction peut résoudre ce problème, avec un expert Planification, vu les déplacements nombreux. Nous adoptons la même hypothèse de compétition entre stratégies que celle des auteurs.

La précédente expérience nous a permis d'évaluer les interactions de stratégies où elles étaient en compétition pendant un essai seulement, celle-ci nous permettra d'estimer les effets d'une répétition de cette mise en situation de conflit.

7.2.1 Protocole et résultats observés

Dans cette expérience, 3 groupes de rats (Contrôle, Hc, Trial) apprennent à trouver l'emplacement d'une plateforme non visible dans une piscine de Morris de 2 m de diamètre. Un amer visible (20 cm) est situé dans la piscine à une distance et orientation constantes de la plateforme (20 cm au Nord). Pour les groupes Contrôle et Hc, au début d'une session, la plateforme et l'amer sont déplacés sur l'un des 8 emplacements prévus (Figure 7.8 a), et sont maintenus à cette position durant 4 essais, au terme desquelles une nouvelle session commence. L'expérience dure 11 sessions. Pour le groupe Trial, l'amer et la plateforme sont déplacés à chaque essai.

Les principaux résultats observés dans cette expérience confirment les hypothèses des auteurs. Ils montrent que :

(i) Les rats Hc et Contrôle peuvent apprendre cette tâche au cours de l'expérience, comme la baisse progressive du temps d'échappement l'indique (Figure 7.8 b).

(ii) Les rats Hc sont plus rapides que les rats Contrôles aux premiers essais de chaque session et plus lents aux quatrième essais de ces mêmes sessions (Figure 7.8 b).

(iii) Les rats Trial ont les mêmes performances que les rats Hc (Figure 7.8 c).

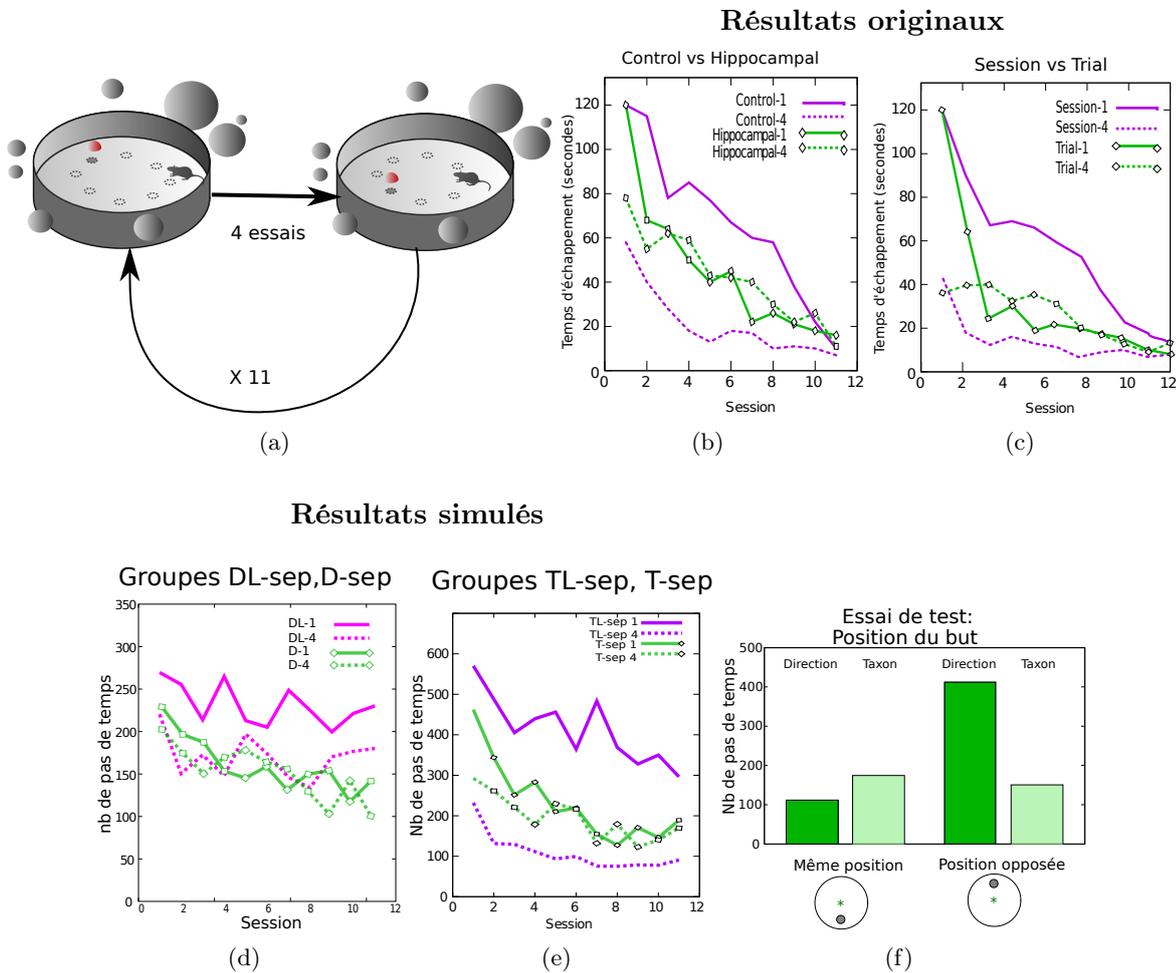


Figure 7.8 – **Expérience de Pearce et al (1998)** (a) L’environnement utilisé en simulation pour reproduire l’expérience de Pearce et al. (1998). (b-c) **Résultats originaux** : apprentissage des animaux Contrôle-Hc et Session-Trial pour les premiers (“-1”) et quatrièmes (“-4”) essais d’une session au cours des sessions. (d-e) **Résultats simulés** : performances obtenues en simulation par les groupes DL-sep/D-sep et TP-sep/T-sep. (f) Simulation du test de vérification du cadre de référence allocentré de l’expert de guidage.

Les auteurs déduisent de ces résultats que les rats Contrôle utilisent deux stratégies pour trouver le but : une stratégie dépendante de l'hippocampe – fondée sur une “carte cognitive” – qui mémorise l'emplacement du but en fonction des indices distaux et une stratégie indépendante de cet hippocampe – fondée sur les “heading-vectors” – qui mémorise l'emplacement du but en fonction de sa direction allocentrique par rapport à l'amer visible.

Ils démontrent l'utilisation d'une telle *stratégie de guidage non égocentrée* par un test consistant à vérifier si les animaux se dirigent vers l'amer, puis cherchent la plateforme au hasard autour de l'amer (stratégie égocentrée), ou bien s'ils ont appris la position du but en allant bien vers le sud de l'amer (stratégie allocentrée). Ils disposent pour cela la plateforme au nord de l'amer et constatent que les animaux se dirigent significativement plus vers le sud que vers le nord.

Ils déduisent aussi que la stratégie hippocampique serait responsable de l'amélioration de la performance des rats intacts au sein d'une session, tandis que la stratégie non hippocampique serait responsable de l'apprentissage de la tâche au cours des sessions. En effet, seule la stratégie basée sur l'amer pourrait trouver la plateforme lorsque celle-ci et l'amer correspondant sont déplacés au début de chaque session. A l'inverse, la stratégie hippocampique induirait la mémorisation du précédent emplacement de la plateforme en mémoire et y mènerait les rats, d'où une hypothèse supposant une compétition entre les deux stratégies aux premiers essais de chaque session.

Le comportement des rats Trial serait semblable à celui des rats Hc, car ils n'auraient pas pu suffisamment construire une représentation spatiale de l'environnement.

7.2.2 Protocole et résultats simulés

La piscine, l'animal et l'amer simulés sont représentés par des cercles de diamètres respectifs 200 cm, 15 cm et 20 cm. La plateforme de 10 cm de diamètre est toujours située 20 cm au Sud de l'amer. Au début de chaque session, la plateforme et l'amer sont déplacés vers un des 8 emplacements possibles (Figure 7.8 a). Au début de chaque essai, l'animat est déposé au bord de la piscine, à l'un des 4 points cardinaux, choisi aléatoirement.

Dans Dollé et al (2008) et Dollé et al (2010a), nous avons simulé notamment ces animaux avec des groupes constitués d'experts Taxon égocentré et Local. Alors que les résultats comprenant l'expert Local ne correspondaient pas aux résultats observés (Figure 7.8 d), ceux comprenant un expert Taxon pouvaient convenir (Figure 7.8 e). Mais ayant reproduit le test de vérification du cadre de référence de Pearce et al (1998) mentionné plus haut, consistant à vérifier si les animaux se dirigeaient bien au sud de l'amer (stratégie allocentrée) ou s'ils cherchaient au hasard autour de l'amer (stratégie égocentrée), il s'est avéré que seuls les animats constitués d'expert Direction, et non Taxon, réussissaient ce test (Figure 7.8 f).

Nous avons donc écarté les groupes avec expert Taxon et ne développons ici que les résultats des groupes utilisant des experts Direction et Planification.

Les groupes suivants sont simulés :

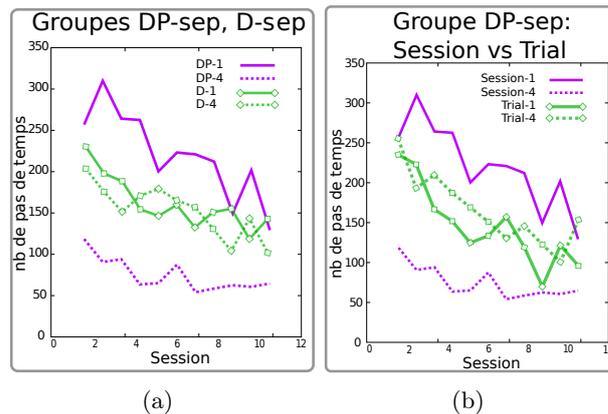


Figure 7.9 – **Simulation de l'expérience de Pearce et al (1998)** (a) Performances obtenues par les groupes DP-sep/D-sep et (b) DP-sep/DP-sep Session vs Trial.

- (i) Les groupes DP (supposés équivalents à Contrôle, Trial) et D (supposés équivalents à Hc)
- (ii) Un groupe P (expert Planification seulement) représentera les animaux lésés du DLS (non inclus dans l'expérience originale) qui seraient censés n'utiliser qu'une stratégie Planification. Il va permettre d'estimer l'influence de la stratégie de lieu.

Ces groupes sont simulés avec les entrées séparées (1 amer proximal pour D, 8 amers distaux pour P) et partagées (9 amers pour D, 9 amers pour P). Pour une cohérence des groupes, les groupes avec entrées séparées (-sep), (accès à la plateforme visible pour expert Direction, aux amers distaux pour expert Planification) seront toujours testés ensemble, ainsi que les groupes avec entrées partagées (-par) (accès à la fois à la plateforme visible et aux amers distaux pour tous les experts).

7.2.2.1 Seuls les groupes DP-sep/D-sep reproduisent la tâche.

Ces groupes reproduisent en effet les comportements mentionnés ci-dessus. (i) Les groupes simulés DP-sep et D-sep apprennent l'emplacement de la plateforme, comme le montre l'évolution de leur temps d'échappement (Figure 7.9 a, $p < 0.001$ pour tous les groupes).

(ii) Contrairement au groupe D-sep, les animaux du groupe DP-sep diminuent significativement leur temps d'échappement pendant les sessions (Control-1 et Control-4, Figure 7.9 a).

(iii) Le comportement du groupe Trial (DP-sep) a les mêmes caractéristiques que le groupe D-sep (Figure 7.9 b).

Les groupes DP-par et D-par ne reproduisent que partiellement la tâche (cf Annexe E pour les Figures). Le groupe DP-par ne montre pas d'apprentissage inter-sessions ($p = 0.09$), tout comme le groupe D-par (Figure E.3 a), ce qui suggère une difficulté pour l'expert Direction d'apprendre correctement l'emplacement de la plateforme en présence d'amers distaux.

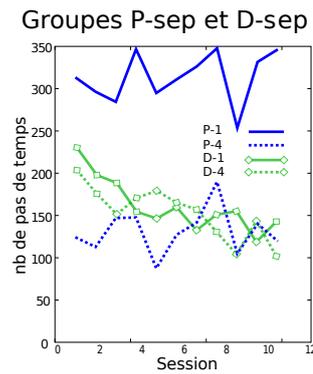


Figure 7.10 – **Simulation de l'expérience de Pearce et al (1998)** Performances obtenues par les groupes D-sep/P-sep.

En revanche, un apprentissage important est observé pendant chaque session, du notamment à l'expert Planification (Figure E.3 b).

Nous nous concentrons pour la suite des analyses sur les groupes avec entrées séparées (DP-sep, D-sep et P-sep).

7.2.2.2 Partage des rôles entre les experts dans le groupe DP-sep

Les hypothèses avancées par les auteurs quant aux rôles des deux stratégies se retrouvent dans les résultats suivants.

L'expert de guidage est responsable de l'apprentissage inter-sessions (1^{ers} essais). La comparaison des deux groupes "lésés" (D-sep et P-sep) montre que l'expert Direction serait responsable de la diminution des temps d'échappement entre les sessions (Figure 7.10). En effet la simulation de l'expert Planification seul ayant construit sa carte avec les amers distaux montre qu'il ne parvient pas apprendre le brusque déplacement de la plateforme, alors que l'expert Direction y parvient, en la positionnant avec l'amer. Le groupe Trial a cette même capacité d'apprentissage inter-sessions, ce qui prouve que l'expert Direction y joue un rôle prépondérant. C'est ce que l'on peut constater dans les sélections d'experts au cours des premiers essais dans la Figure 7.11 a, où l'expert Direction prend progressivement le contrôle sur l'expert Planification—plus tôt dans le groupe Trial (Figure 7.11 c, gauche).

L'expert de lieu est responsable de l'apprentissage intra-session. La comparaison des mêmes groupes lésés (D-sep et P-sep) montre que l'expert Planification serait responsable de la diminution des temps d'échappement pendant une session (Figure 7.9 b comparée à 7.10).

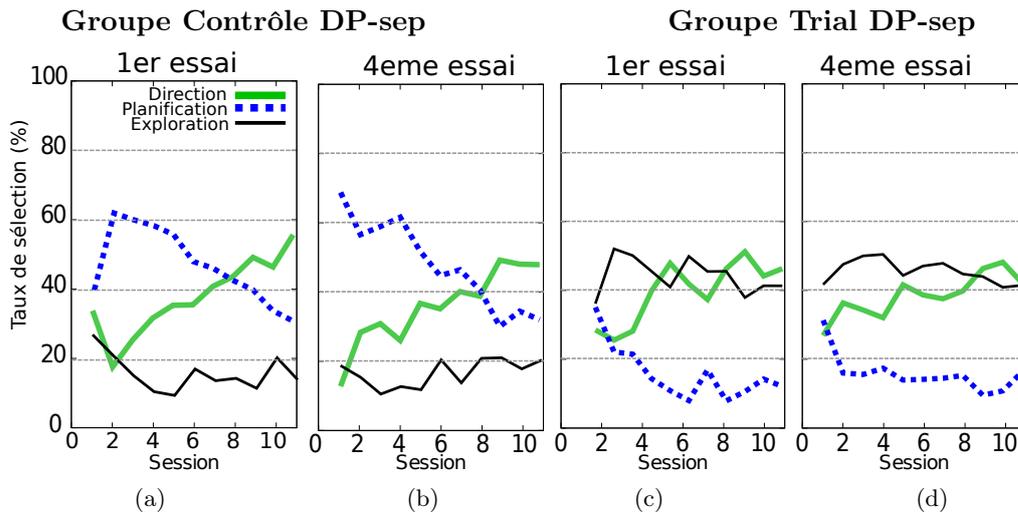


Figure 7.11 – **Simulation de l'expérience de Pearce et al (1998)** (a-b) Groupe Contrôle (=Session) DP-sep; (c-d) Groupe Trial DP-sep : taux de sélection par session des experts dans les premiers et quatrièmes essais.

7.2.2.3 L'apprentissage du groupe Contrôle DP-sep s'explique par les interactions entre ces deux experts

Conformément aux hypothèses des auteurs, la moins bonne performance inter-sessions du groupe Contrôle (1^{ers} essais) s'explique par une compétition d'experts de guidage et de lieu. Le groupe avec experts Direction et Planification rejoint la plateforme moins rapidement aux 1^{ers} essais que les groupes avec expert Direction seulement (Figure 7.9 c), suggérant l'existence d'une compétition entre les stratégies de guidage et de lieu, cette dernière ralentissant l'apprentissage de l'autre. En comparant les taux d'occupation du but cumulés pendant les premiers essais entre les groupes DP-sep, D-sep et P-sep (Figure 7.12 a,b), on constate que le groupe DP-sep a le même taux d'occupation du but à l'emplacement courant que le groupe D, mais, qu'en revanche, il a un plus fort taux d'occupation du but précédent que ce groupe D.

Le groupe Planification a, lui, un taux d'occupation du but précédent encore plus important (Figure 7.12 b), d'où l'hypothèse (déjà évoquée dans le test de compétition de Devan et White, 1999) que cet expert entraînerait les animats vers le but précédent, pendant que l'expert Direction les mènerait vers la plateforme. Une trajectoire typique de cette situation est illustrée Figure 7.13 a : l'expert Planification guide l'animat vers le précédent emplacement de la plateforme (cf. carte de l'expert Planification pointant sur l'emplacement précédent : Figure 7.13 b), tandis que l'expert Direction le mène finalement vers son emplacement courant (Figure 7.13 a, carte en c).

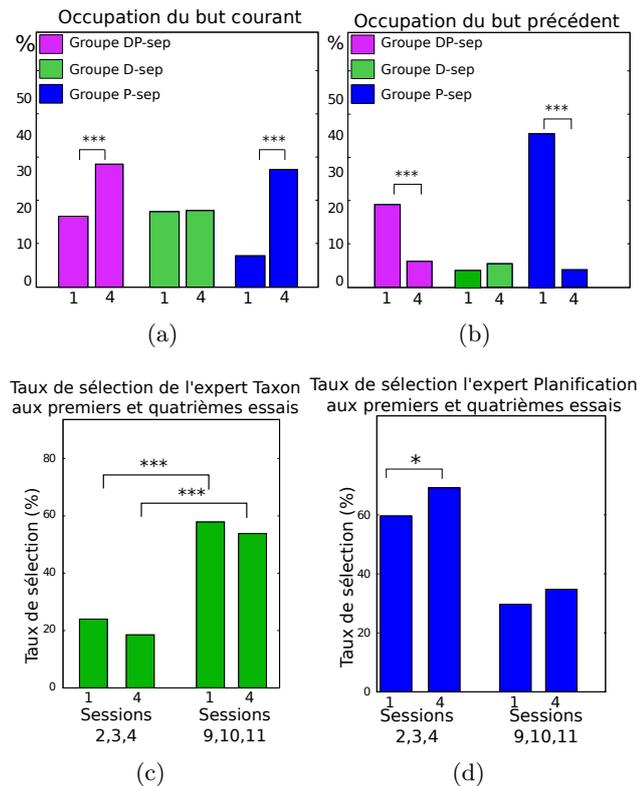


Figure 7.12 – **Simulation de l'expérience de Pearce et al (1998)** (a-b) Groupes DP-sep, D-sep et P-sep : taux d'occupation du but à l'emplacement courant et précédent *** correspond à $p < 0.001$; (c) Taux de sélection au but de l'expert Planification et (d) de l'expert Direction entre les premiers et quatrièmes essais dans les premières et dernières sessions.

La compétition diminue au fur et à mesure de l'expérience. Les performances du groupe DP-sep rejoignent celles du groupe D-sep en fin d'expérience. la Figure 7.11 b montre en effet la prise de contrôle finale de l'expert Direction sur l'expert Planification. La compétition entre les deux experts diminue au profit de l'expert de guidage. Ceci est confirmé par la Figure 7.12 c,d, qui indique les taux de sélection des experts en début et en fin d'expérience : celui de l'expert de guidage augmente, alors que celui de l'expert de lieu diminue, des premières aux dernières sessions. L'expert Direction a eu en effet le temps d'apprendre à préciser la position du but, comme l'illustration de sa carte navigationnelle le démontre (Figure 7.13 c).

A la différence de l'hypothèse des auteurs, la meilleure performance intra-session du groupe Contrôle s'expliquerait par une sélection conjointe d'experts de guidage et de lieu. Le groupe DP-sep rejoint la plateforme plus rapidement aux 4^{èmes} essais que le groupe avec expert Planification seulement (Figure 7.9 c comparée à 7.10), suggérant l'existence d'une coopération entre les stratégies de guidage et de lieu. En comparant les taux d'occupation du but cumulés des experts pendant les quatrièmes essais entre les groupes DP, D et P (Figure

7.12 a, b), on remarque que l'occupation du but courant est la même dans les groupes DP et P, alors qu'elle est inchangée dans le groupe D. On pourrait conclure que seul l'expert Planification contribue à cette amélioration – ce qui était d'ailleurs l'hypothèse des auteurs – pourtant les sélections d'experts au cours des 4^{èmes} essais (Figure 7.11 b) ne la confirment pas. L'expert Planification y est sélectionné plus que pendant les premiers essais, mais l'expert Direction est sélectionné aussi et prend même le contrôle à la fin de l'expérience. Ceci est confirmé par la Figure 7.12 c, d : les taux de sélections aux quatrièmes essais sont plus importants pour l'expert de lieu et se maintiennent pour l'expert Direction.

Cette meilleure performance serait due à une coopération entre experts. Malgré la similitude des sélections d'experts aux premiers et quatrièmes essais, le rôle de l'expert Planification est inverse par rapport à celui observé aux premiers essais : ici, il coopérerait avec l'expert Direction pour positionner le but. Une telle coopération est illustrée par une trajectoire typique en milieu d'expérience (session 6), qui montre que l'expert Planification est plus souvent sélectionné au début, puis l'expert Direction ensuite (Figure 7.13 d). Vers la fin de l'expérience (session 11), l'expert Direction est plus présent en début de trajectoire car il est devenu plus efficace que l'expert Planification (Figure 7.13 e). Des exemples de cartes de navigation correspondant à la trajectoire e (Figure 7.13 f, g) montrent que l'expert de guidage localise toujours l'emplacement de la plateforme par rapport à l'amer et que l'expert Planification mémorise maintenant l'emplacement courant et non le précédent.

7.2.2.4 Groupes Trial DP-sep

Le groupe Trial DP-sep se comporte comme un groupe D-sep. On constate en effet que le groupe Trial (Figure 7.9 e), dont l'expert Planification peut intégrer les amers distaux, présente les mêmes performances qu'un groupe D-sep, dont l'expert de guidage ne s'ancre que sur l'amer proximal. Les taux de sélection montrent en effet que l'expert Direction est sélectionné progressivement avec beaucoup plus d'exploration aléatoire que dans les groupes DP-Session et que l'expert Planification est progressivement dévalué (Figure 7.11 c). Les sélections ne changent pas au cours d'une session, ce qui se traduit aussi dans les performances.

L'hypothèse des auteurs quant à la nécessité de rafraîchir une carte cognitive en cas de changements trop rapides d'emplacements de la plateforme est ici avérée.

7.2.2.5 Autres groupes

La compétition entre les experts Direction diminue les performances du groupe DP-par (cf Annexe E pour les Figures). Les performances de ces groupes sont en effet moins bonnes que celles des groupes DP-sep et D-sep. On y constate une légère amélioration de l'apprentissage pendant les premiers essais dans le groupe DP comparé à P, mais pas d'amélioration dans les quatrièmes essais (Figure E.3 a, b). Les taux de sélection des experts indiquent une baisse de sélection pour tous les experts, sans augmentation nette de l'un d'eux aux premiers et derniers essais (Figure E.4 a, b).

Le groupe D-par a un moins bon apprentissage inter-sessions que le groupe D-sep. La Figure E.4 c, d montre que l'expert Direction ancré sur l'amer proximal contrôle

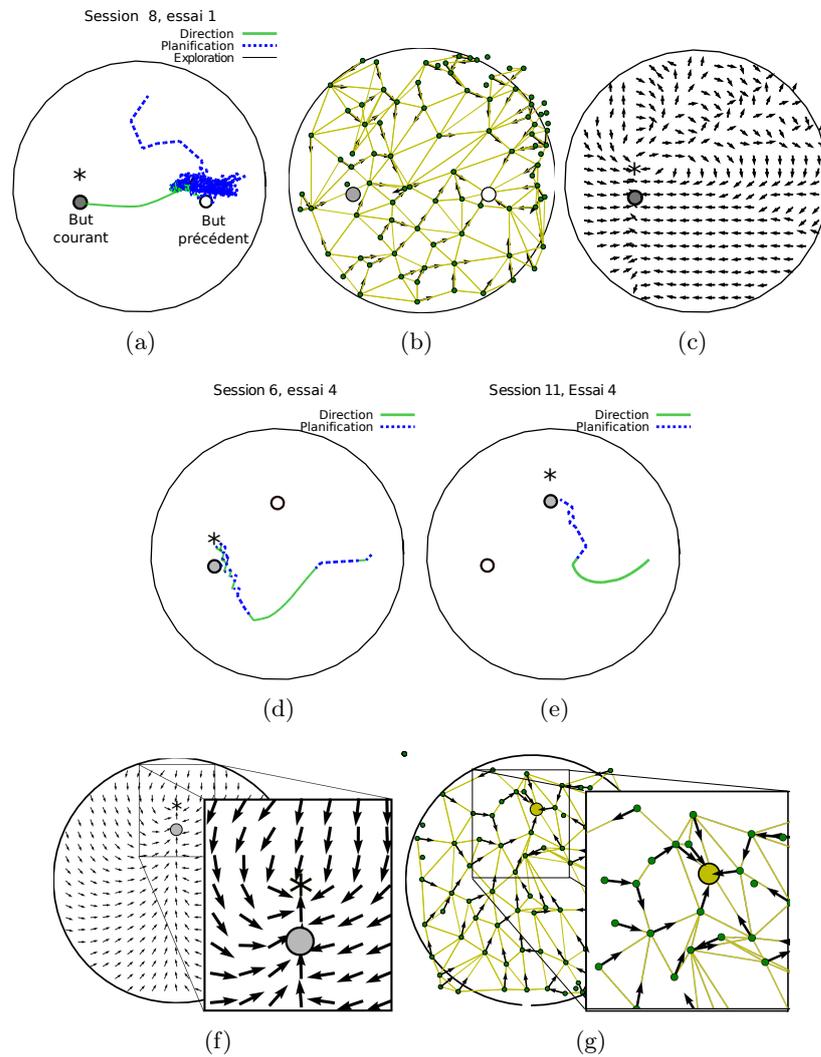


Figure 7.13 – **Simulation de l'expérience de Pearce et al (1998)** Groupe DP-sep (a) Trajectoire typique durant le premier essai de la session 8 ; (b, c) Cartes navigationnelles des experts Direction et Planification correspondant à la trajectoire a. (d, e) Trajectoires typiques au début (session 4) et à la fin (session 11) ; (f, g) : Cartes navigationnelles des experts Direction et Planification correspondant à la trajectoire e. Sur ces cartes, les flèches correspondent aux mouvements en ce point de l'environnement (expert Direction) ou à un nœud du graphe (expert Planification). L'astérisque correspond à la position courante de l'amer, les cercles pleins et évidés correspondent respectivement aux emplacement courant et précédent du but.

majoritairement le comportement dès la quatrième session, mais la sélection des experts ancrés sur les amers distaux perturbe l'apprentissage plutôt qu'il l'améliore – jouant un rôle semblable à celui de l'expert Planification dans le groupe DP-sep. Ces experts sont dévalués au cours d'une session (Figure E.4 d), au profit de l'exploration aléatoire. Cette exploration semble également jouer le rôle de l'expert Planification dans le groupe DP-sep.

7.2.3 Discussion partielle

Conformément aux hypothèses de Pearce et al (1998), les animats ont appris parallèlement une stratégie de guidage et une stratégie de lieu dans une tâche dans laquelle une plateforme indicée est déplacée tous les 4 essais. Ces deux stratégies (Direction et Planification) ne sont pas les mêmes que dans la situation expérimentale de Devan et White (1999) (Taxon et Local), mais présentent le même type de compétition pendant l'essai où la plateforme est déplacée : la stratégie de lieu propose une orientation différente de celle de la stratégie de guidage. Cette compétition peut diminuer au fur et à mesure de l'expérience, comme la condition de Pearce et al (1998) le démontre, la stratégie de lieu perdant progressivement sa capacité de perturber l'orientation correcte des animats. L'interaction entre les experts a pu rendre compte de ces deux phénomènes : compétition et diminution de la compétition.

Un apprentissage intra-session conséquent est dû à la stratégie de lieu. Alors qu'en début de session nous avons confirmé les interactions compétitives entre stratégies, nous suggérons qu'au cours de chaque session les deux stratégies coopèrent pour mieux positionner la plateforme. La carte cognitive de l'expert de lieu pointe rapidement au sud de l'amer, alors que l'expert Direction apprend encore à localiser le but. La rapidité de l'apprentissage de la stratégie de lieu avait déjà été démontrée dans la simulation de l'expérience de Steele et Morris (1999). L'expert Planification parvient à améliorer ses performances dès le deuxième essai, et l'expert Direction peut encore améliorer cet apprentissage. La coopération s'effectuerait de façon flexible, elle pourrait alterner l'utilisation de stratégies de lieu et/ou de guidage selon le niveau de compétence acquis par chacune des stratégies.

Contrairement à la situation expérimentale de Devan et White (1999) où les deux stratégies Taxon et Locale étaient également compétentes pour positionner un but fixe, ici les déplacements fréquents du but (une situation de conflit répétée) ont donc déclenché une coopération entre les deux experts, l'un (Planification) étant plus efficace que l'autre (Direction) pour atteindre rapidement la plateforme pendant les quelques essais de chaque session, l'autre (Direction) pouvant conduire plus efficacement les animats au début de chaque session et en fin d'expérience. Une telle interaction pourrait donc être envisageable, non pas entre DLS et Hc (hypothèse de Devan et White (1999) évoquée plus haut), mais entre DLS et mPFC.

L'hypothèse de systèmes parallèles est aussi étayée par la similitude de performance d'animats Contrôle ne pouvant pas s'ancrer durablement sur les amers distaux et d'animats sans expert de Planification, comme l'avait observé les auteurs entre groupes d'animaux Trial et Hc. Ces deux groupes présentaient tous deux des performances moins bonnes qu'un groupe pouvant s'entraîner pendant quelques essais d'une session (groupe Session). On a constaté les mêmes types

de différences entre les groupes Session et Trial de Roberts et Pearce (1999) et les groupes Session et Trial de Pearce et al (1998), bien que les conditions expérimentales aient été différentes : chez Roberts et Pearce (1999), l'amer était situé sur le but et non indicé à côté (Figure 6.11). Bien qu'aboutissant à des effets semblables, les courbes d'apprentissage du groupe Session de Pearce et al (1998) sont le résultat d'une sélection d'experts très différente de celle observée dans le groupe Session de Roberts et Pearce (1999) (Figure 6.10 b comparée à 7.12 c,d) : la sélection de l'expert Direction ancré sur l'amer proximal est très diminuée et l'exploration y prend beaucoup d'importance. Des différences de performance s'ensuivent : dans la simulation de Pearce et al (1998), dues à une alternance plus fréquente de sélection entre les experts et l'exploration, les performances des groupes Session et Trial ne deviennent (presque) semblables qu'au bout de 12 sessions, au lieu de 5 seulement dans la simulation de Roberts et Pearce (1999).

Il est à remarquer que l'on observe le même effet chez les animaux : dans l'expérience de Roberts et Pearce (1999) (Figure 6.6 b), ils obtiennent les mêmes performances au bout de 2 ou 3 sessions, au lieu de 12 dans l'expérience de Pearce et al (1998) (Figure 7.9 a). La position de l'amer proximal vis-à-vis du but a donc un impact non négligeable dans l'efficacité d'une stratégie de guidage à compenser la non-accessibilité aux amers distaux.

La distance entre l'amer proximal et le but n'est pas le seul facteur à influencer les interactions de stratégies. Ce que nous avons déjà relevé est leur possibilité d'entraînement, révélée ici par la différence entre les groupes Session et Trial, le premier permettant la prise en compte des amers distaux par la pratique de seulement 4 essais successifs par session (cf aussi Dollé et al, 2010a, pour une analyse plus détaillée).

7.3 Conclusion du chapitre

Plus que dans les autres expériences, on aura sûrement remarqué la difficulté avec laquelle les résultats simulés ont reproduit dans ce chapitre les résultats observés (résumés dans le tableau 7.1). Ces simulations ont été en effet particulièrement sensibles aux paramètres du modèle (e.g., taux d'apprentissage des experts, du réseau de sélection, etc.). Cela tient au fait qu'il a fallu ici retrouver précisément des performances au fur et à mesure de l'apprentissage, et non juste des effets globaux comme des taux d'occupation de telle ou telle zone de la piscine lors d'un essai test.

La reproduction de ces deux expériences de groupes intacts et lésés, mettant explicitement en situation de conflit deux stratégies de lieu et de guidage – respectivement construisant ou non une représentation spatiale –, a permis d'aller dans le sens des hypothèses des auteurs quant à l'existence de systèmes de navigation apprenant en parallèle et recrutant spécifiquement, l'un les amers proximaux, l'autre les amers distaux. En effet, lorsque les experts du modèle ont accès à tous les amers, les résultats obtenus n'ont pas pu rendre aussi bien compte du comportement des animaux lorsqu'ils sont mis dans une situation de conflit, lorsqu'un amer proximal leur indique plus que les amers distaux le changement d'emplacement de la plateforme.

Interactions contextuelles d'experts en situations de conflit. Des situations de conflit telles que celles proposées dans ces deux expériences, où l'amer proximal est toujours mis en position d'indiquer plus efficacement la plateforme, aboutit de façon logique à une prise de contrôle rapide d'une stratégie de guidage. Cette prise de contrôle est cependant très dépendante du contexte environnemental. En effet, chez Devan et White (1999) où la plateforme est visible pendant les premières sessions, la stratégie de guidage prend le contrôle immédiatement, alors que c'est la stratégie de lieu qui est immédiatement majoritaire lorsque l'amer proximal est situé à côté du but, comme chez Pearce et al (1998). Il s'ensuit, dans cette dernière situation, une marge de progression de la stratégie de guidage qui a été essentielle pour mettre en évidence la compétition graduelle des deux systèmes de guidage et de lieu au cours de l'expérience. La Figure 7.14 a montre en effet la courbe générale (tous essais confondus) d'apprentissage, plus lente que celle d'un même groupe Contrôle (DP-sep) de Devan et White (1999), ainsi que le nombre de sessions que met la stratégie de guidage pour prendre le contrôle majoritairement. Il faut en effet au moins 8 sessions pour que la sélection de l'expert de guidage soit majoritaire chez Pearce et al (1998) (Figure 7.14 b). Il n'en faut que 2 chez Devan et White (1999) (Figure 7.14 c).

Malgré ces situations expérimentales conflictuelles, on a vu que des interactions de coopération pouvaient s'intercaler entre des interactions de conflit. Cela pourrait indiquer que chaque système apprend et exécute en parallèle une stratégie, qui pourrait perturber ou améliorer de manière "passive" le comportement. Cela ne semble pas être le cas. Les interactions, notamment de coopération, semblent plus subtiles, car elles n'améliorent le comportement *que si* les deux stratégies sont réellement complémentaires (comme dans Pearce et al, 1998) – et non pas toutes deux aussi efficaces (comme dans Devan et White, 1999).

Effets associatifs et cognitifs Dans ces deux expériences se retrouve la nécessité de prendre en compte une compétition entre amers, non pas entre tous les amers, mais plutôt entre systèmes focalisant respectivement sur les amers proximaux et distaux. En effet, comme mentionné plus haut, les comportements correspondants ont pu difficilement émerger par une compétition entre tous les amers, les résultats des groupes à entrées partagées n'ayant pas été très probants.

Des effets d'occultation entre amers proximaux et distaux ont pu toutefois apparaître dans les deux expériences, dans lesquelles ces deux systèmes apprennent de façon simultanée. Dans l'expérience de Devan et White (1999), une diminution de la saillance de l'amer proximal par les amers distaux, suite à quelques sessions pendant lesquelles il était absent, a eu pour effet de diminuer les performances du groupe Contrôle à l'essai de compétition. Dans l'expérience de Pearce et al (1998) au contraire, une diminution progressive de la salience des amers distaux par l'amer proximal pendant les premiers essais de chaque session a permis d'augmenter les performances du groupe Contrôle.

Est-il nécessaire, dans ces expériences, de faire intervenir un expert déclaratif pour représenter la stratégie de lieu ? Dans la reproduction de l'expérience de Pearce et al (1998), la présence d'un expert Planification s'est avérée essentielle pour expliquer la rapidité d'apprentissage au cours

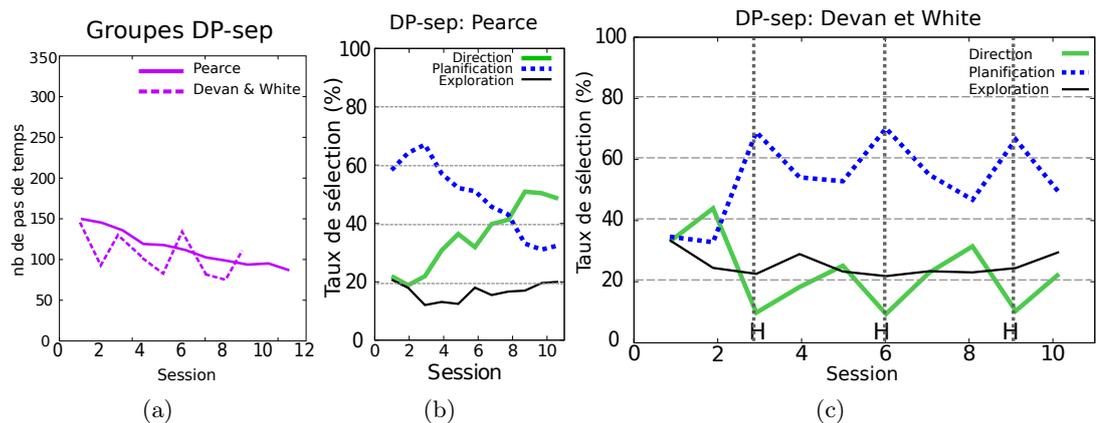


Figure 7.14 – **Simulations de Pearce et al (1998) et Devan et White (1999)** (a) Courbes d'apprentissage générale des groupes DP-sep dans les deux expériences. (b) Comparaison de la sélection d'experts chez Pearce et al (1998) et (c) chez Devan et White (1999) pour les mêmes groupes DP-sep.

d'une session. On a vu qu'un expert Local n'avait pas la rapidité d'apprentissage intra-session nécessaire à la bonne reproduction du comportement des groupes Contrôle (voir Tableau 7.1).

Dans la reproduction de l'expérience de Devan et White (1999), le cas est moins évident. Un expert Local qui s'ancre sur les amers distaux, sans pour cela reproduire parfaitement les performances des animaux, semble avoir une vitesse d'apprentissage plus conforme à celle des animaux DLS. Une association procédurale entre actions, *mais avec carte hippocampique intégrant les amers distaux* pourrait donc suffire. Cela pourrait s'expliquer par la position fixe du but, qui inciterait les animaux à opter pour une stratégie moins coûteuse qu'une stratégie cognitive (voir Tableau 7.1).

La partie suivante présente un bref rappel des résultats obtenus, puis une discussion générale sur les contributions et limitations du modèle aux hypothèses relatives à l'apprentissage spatial.

Tableau 7.1 – Résumé des résultats simulés des groupes Contrôles de ce chapitre, selon les configurations d’amers et si le but est déplacé ou non pendant l’entraînement (Ent.) et les test (Test). Les expériences sont indiquées par le nom du premier auteur de l’article correspondant. Les groupes "Conforme" sont ceux qui reproduisent au plus près le comportement animal; les groupes "Possible" sont ceux qui présentent quelques différences avec le comportement observé; les "Non conforme" ne reproduisent pas les résultats. Les groupes "Non simulé" n’ont pas participé à ces expériences. Enfin les sous-catégories "Isol.", "Sep." et "Par." signifient si les simulations ont été réalisées avec, respectivement, un experts seul, des experts avec entrées séparées ou partagées.

Exp.	Amers	But déplacé		Conforme			Non conforme		
		Ent.	Test	Isol.	Sep.	Par.	Isol.	Sep.	Par.
Devan	1 balise 8 dist	non	oui		TL			DP	DP
Pearce	1 prox 8 dist	oui	oui		DP			DL,TP	DP

Exp.	Amers	But déplacé		Possible			Non simulé		
		Ent.	Test	Isol.	Sep.	Par.	Isol.	Sep.	Par.
Devan	1 balise 8 dist	non	oui		TP	TL		DL	DL
Pearce	1 prox 8 dist	oui	oui					TL	TP,DL

Cinquième partie

Discussion et conclusion

Chapitre 8

Discussion générale

Le modèle computationnel de sélection de stratégies de navigation présenté dans ce travail a analysé diverses influences sur l'apprentissage spatial de repères visuels environnementaux, plus ou moins nombreux et plus ou moins proches de la zone de récompense (amers proximaux et distaux). Cet apprentissage serait effectué, pour les uns, en mettant en compétition tous les amers (théorie associative), pour d'autres, en intégrant tous les amers dans une représentation cohérente de l'environnement (théorie cognitive). Une conception plus consensuelle serait que ces deux types d'apprentissages cohabiteraient dans des modules parallèles, pouvant avoir des interactions compétitives ou synergiques. Le modèle bio-inspiré vise à contribuer au débat concernant l'apprentissage de ces amers en faisant l'hypothèse que ces interactions peuvent expliquer certains effets avancés par l'un ou l'autre des deux points de vue antagonistes.

Outre des simulations réalisées avec la présence d'un seul type d'amers (proximaux ou distaux) destinées à évaluer le comportement du modèle (EXP1, EXP2, Morris et al, 1982; Steele et Morris, 1999), des expériences permettant une validation de cette contribution ont été choisies pour l'existence ou l'absence d'effets compétitifs, voire coopératifs, entre amers proximaux et distaux. Les expériences de Rodrigo et al (2006) et de Roberts et Pearce (1999) mettent en évidence des effets compétitifs par l'influence de la distance entre amers ou par le blocage de l'apprentissage d'amers distaux par celui d'amers proximaux, alors que l'expérience de McGregor et al (2004a) démontre l'absence d'effet compétitif dans leurs apprentissages. De même, émergeant à la conception des systèmes modulaires, les expériences de Devan et White (1999) et Pearce et al (1998) démontrent une compétition entre systèmes, en mettant explicitement en conflit ces deux types d'amers.

Les résultats des simulations des groupes Contrôles – censés pouvoir utiliser à la fois une stratégie de guidage et une stratégie de lieu –, sont rassemblés dans le Tableau 8.1, selon l'ordre dans lequel nous les avons présentés dans les chapitres précédents. Ces simulations ont été réalisées avec des instanciations différentes selon les expériences (experts différents) qui toutes suivent les mêmes principes essentiels de la modélisation.

Dans les sections suivantes nous rappellerons d'abord ces principes de modélisation qui ont permis de reproduire les résultats acquis, puis les contributions essentielles de ces résultats aux hypothèses relatives à l'apprentissage spatial. Les facteurs qui ont pu en limiter la portée, comme

certain aspects de la conception du modèle ou les contraintes de tâches imposées dans le cadre de ces simulations, seront ensuite passés en revue. Dans la dernière section seront exposées les raisons pour lesquelles le modèle peut être un outil d'aide à l'analyse et à la prédiction de résultats d'expérimentations animales sur la navigation.

8.1 Principes essentiels de l'architecture du modèle

Cette section rappelle en quoi la conception du modèle peut être spécifiquement adaptée au test d'hypothèses relatives à l'apprentissage spatial.

Experts séparés Le modèle est modulaire, construit sur la base d'experts de navigation séparés. Ces experts ont été conçus au départ pour accepter en entrées des amers spécifiques, soit les amers proximaux pour les experts de guidage, les amers distaux pour les experts de lieu (groupes "-sep"). Cependant rien dans le modèle (comme chez l'animal) n'empêche les experts d'apprendre sur la base de tous les amers, ce que nous avons effectué dans toutes les simulations (groupes "-par"). Cette possibilité permet de tester les hypothèses associatives, cognitives et modulaires.

L'apprentissage d'un expert de guidage procédural ayant accès à tous les amers (groupes T-par ou D-par) peut en effet tester l'hypothèse associative. Les amers sont ainsi indépendamment en compétition et cet expert devrait, d'après cette théorie, générer des phénomènes de blocage et occultation. L'apprentissage d'un expert de lieu, intégrant à la fois les amers proximaux et distaux dans sa représentation spatiale (groupes L-par ou P-par), peut tester l'hypothèse cognitive. Tous les amers contribuent à la recherche du but et cet expert devrait, d'après cette théorie, générer une absence d'effets associatifs. Les interactions d'experts avec entrées partagées ou séparées peuvent tester de même l'hypothèse consensuelle, qui serait que certains amers, pris en entrée de façon indépendante mais aussi de façon intégrée, peuvent rendre compte de résultats contradictoires. Si l'aspect modulaire était déjà présent dans certains modèles computationnels (e.g., Guazzelli et al, 1998; Girard et al, 2005; Chavarriaga et al, 2005), les possibilités de prendre en compte ainsi certains ou tous les amers n'ont pas – ou pas pu – être exploitées.

Apprentissages différentiels. Des mécanismes d'apprentissage d'associations entre amers et actions, dépendant ou non d'une récompense (procédural et déclaratif), sont implantés dans des modules séparés. Ils ont chacun leurs caractéristiques, le premier étant plus lent, moins flexible, mais plus performant que le deuxième après un entraînement plus long. Ils peuvent donc permettre, d'une part, de vérifier si l'apprentissage spatial est sous la dépendance d'un même ou de différents mécanismes d'acquisition et, d'autre part, de préciser des interactions de stratégies supposées dépendantes de ces deux modes. D'autres modèles computationnels existants proposent ces deux types d'apprentissage (e.g., Guazzelli et al, 1998; Girard et al, 2005), mais sans avoir la possibilité de confronter véritablement leurs performances.

Apprentissages réciproques. Nous avons mentionné dans le Chapitre 4 que l'apprentissage parallèle des experts pendant l'expérience est conçu de telle sorte que l'apprentissage de l'un peut influencer l'apprentissage de l'autre, *même si celui-ci n'est pas sélectionné*. D'après certains

Tableau 8.1 – Résumé des résultats simulés des groupes Contrôles, selon les configurations d’amers et si le but est déplacé ou non pendant l’entraînement (Ent.) et les test (Test). Les expériences sont indiquées par le nom du premier auteur de l’article correspondant. Les groupes "Conforme" sont ceux qui reproduisent au plus près le comportement animal (ou, dans le cas d’EXP1, EXP2, ceux qui sont les plus performants) ; les groupes "Possible" sont ceux qui présentent quelques différences avec le comportement observé ; les "Non conforme" ne reproduisent pas les résultats. Les groupes "Non simulé" n’ont pas participé à ces expériences. Enfin les sous-catégories "Isol.", "Sep." et "Par." signifient si les simulations ont été réalisées avec, respectivement, un expert seul, des experts avec entrées séparées ou partagées.

Exp.	Amers	But déplacé		Conforme			Non conforme		
		Ent.	Test	Isol.	Sep.	Par.	Isol.	Sep.	Par.
A- Exp1	Balise	non		T D	DP				
B- Exp2	1 prox	non				TP,DP	T	TP	
C- Morris	8 dist	non			DL	DL	P D	DP	DP
D- Steele	8 dist	oui			DP	DP	D	DL	DL
E- Rodrigo	1 balise 1 dist	non	oui		DP		T,D	TP	TD,TP
F- Roberts	1 balise 8 dist	oui	non		DP		T D	TP,DL	TP
G- McGregor (Stable et Rotate)	2 prox 8 dist	non	oui			DP	T,P	TP	TP
H- Devan	1 balise 8 dist	non	oui		TL			DP	DP
I- Pearce	1 prox 8 dist	oui	oui		DP			DL,TP	DP

Exp.	Amers	But déplacé		Possible			Non simulé		
		Ent.	Test	Isol.	Sep.	Par.	Isol.	Sep.	Par.
A- Exp1	Balise	non			TP	TP,DP	L P	TL,DL	TL,DL
B- Exp2	1 prox	non		D	DP	TP	L P	TL,DL	TL,DL
C- Morris	8 dist	non		L			T	TL,TP	TL,TP
D- Steele	8 dist	oui					T L P	TL,TP	TL,TP
E- Rodrigo	1 balise 1 dist	non	oui				L	TL,DL	TL,DL
F- Roberts	1 balise 8 dist	oui	non			DP	L	TL	TL
G- McGregor (Stable et Rotate)	2 prox 8 dist	non	oui	D	DP,DL		L	TL	TL,DL
H- Devan	1 balise 8 dist	non	oui		TP	TL		DL	DL
I- Pearce	1 prox 8 dist	oui	oui					TL	TP,DL

auteurs (e.g., White, 2005; Chang et Gold, 2003), cette influence est possible, puisque ces systèmes sont tous interconnectés via le cortex, et supportée par les observations de changements entre stratégies déclaratives et procédurales observées dans les labyrinthes en T (e.g., Packard et McGaugh, 1996), supposant que les deux stratégies sont apprises simultanément dès le début de l'expérience. Cet aspect du modèle, déjà présent dans le modèle de Chavarriaga et al (2005), s'avère important pour inciter des interactions coopératives entre modules.

Monnaie commune pour la sélection. La proposition de modules apprenant en parallèle avec des mécanismes différents n'aurait pas été opérationnelle sans la recherche d'un critère permettant de comparer leurs performances. La "monnaie" commune à tous les modules considérés est la plus générique possible, puisqu'il s'agit de l'orientation proposée par chacun d'entre eux. Ce critère permet la comparaison des performances de toutes stratégies, qu'elles soient déterministes, dépendant ou non d'une prédiction de récompense, ou aléatoires. Les modèles computationnels confrontant ces différents types de stratégies (e.g., Guazzelli et al, 1998; Girard et al, 2005) ne les sélectionnent que par des systèmes ad hoc.

Sélection adaptative. Le module de sélection, conçu séparément des experts de navigation, peut apprendre de façon adaptative, sur la base de cette monnaie commune, de choisir à chaque pas de temps la stratégie la plus pertinente dans le contexte environnemental courant. Son apprentissage par renforcement de type Q-learning prend en compte à la fois les entrées de chacun des experts et leurs propositions d'orientation. La sélection qui en résulte est donc influencée à la fois par le contexte et par l'historique des choix de stratégies. De par son mode d'apprentissage, elle peut aussi favoriser la compétition entre experts. Les modèles proposant des mécanismes de sélection adaptatifs (e.g., Chavarriaga et al, 2005; Daw et al, 2005) ne pouvaient pas être appliqués à des choix d'experts si variés, ou à des environnements continus.

Les diverses instanciations du modèle, conçues pour reproduire et analyser le comportement spatial du rat dans les quelques expériences retenues, répondent toutes aux mêmes principes de modélisation mentionnés ci-dessus. La section suivante développe quelques exemples du rôle que ces principes ont tenus dans l'acquisition des résultats.

8.2 Contribution des principes de la modélisation aux résultats acquis

8.2.1 Des experts variés et complémentaires

Chaque expert a une compétence qui s'est révélée bien spécifique. L'expert Taxon est plus performant que l'expert Direction dans le positionnement d'un but visible ou d'une balise (Exp1, Tableau 8.1) et l'expert Direction plus performant dans le positionnement du but invisible par un amer proximal (Exp2). L'expert Local est plus performant que l'expert Planification et que l'expert Direction pour positionner un but invisible et fixe par les amers distaux (Morris et al, 1982). L'expert Planification est plus rapide pour repositionner un but régulièrement déplacé que l'expert Local, Direction, et Taxon (Steele et Morris, 1999). La complémentarité de compétences de ces experts a contribué à étayer la représentation modulaire de l'architecture de

navigation. L'absence d'effets associatifs provient de la nécessité pour l'animat d'utiliser des stratégies complémentaires pour résoudre correctement la tâche.

Cette complémentarité a montré par exemple qu'une stratégie de guidage, ou Locale, pouvait être utilisée à la place d'une stratégie de lieu déclarative dans des environnements très stables (Morris et al, 1982; Devan et White, 1999; McGregor et al, 2004a). Cela supporte les hypothèses selon lesquelles l'utilisation d'une "carte cognitive" (carte topologique) n'est pas systématique, même lorsque le but est caché. Bennett (1996) remet ainsi en cause l'existence d'une telle carte dans nombre d'expériences, où, selon lui, des stratégies procédurales peuvent expliquer le comportement des animaux. Hamilton et al (2009), mais aussi Skinner et al (2003) montrent également au moyen d'expériences comportementales simples que, lorsqu'ils ont le choix, les animaux préfèrent utiliser une stratégie procédurale pour rejoindre un but caché selon les amers distaux.

L'utilisation d'experts de guidage fondés sur les indices proximaux et de lieu fondés sur les indices distaux s'est en revanche avérée indispensable pour gérer des conflits entre amers, tels que dans les expériences de Pearce et al (1998) et Devan et White (1999) (Tableau 8.1, I, H).

L'exploration, que nous avons peu évoquée dans l'analyse des résultats, n'en a pas moins contribué étroitement au succès des simulations. Bien que peu réaliste sous la forme avec laquelle nous l'avons simulée (aléatoire), elle a permis de soutenir certains experts à des moments spécifiques. La particularité de cet "expert" est qu'il est ici implémenté comme une stratégie à part, sélectionnée comme les autres par le réseau de sélection. De fait, le besoin d'explorer de nouvelles actions est appris durant l'entraînement et peut dépendre des entrées sensorielles. Cette vision contraste avec l'utilisation de l'exploration dans les algorithmes d'apprentissage par renforcement, où l'exploration est choisie selon un schéma stochastique prédéfini : par exemple, Arleo et Gerstner (2000) utilisent la méthode ϵ -greedy, dans lesquelles une action aléatoire pouvait être choisie avec une faible probabilité, tandis que Foster et al (2000) utilisent une sélection *soft-max*, dans laquelle les actions ont une probabilité de sélection dépendante de la Q-valeur associée. Nous montrons ici que le modèle, dans lequel la balance exploitation/exploration n'est pas prédéfinie mais apprise avec l'entraînement, reproduit efficacement le comportement des animaux. En accord avec les algorithmes standards d'apprentissage par renforcement, l'exploration est choisie majoritairement en début d'expérience puis diminue au fur et à mesure que les autres stratégies sont apprises.

8.2.2 Apprentissages réciproques

L'apprentissage réciproque entre experts a contribué par exemple à la réussite de la tâche par le groupe Contrôle dans l'expérience de Devan et White (1999). Pour estimer les conséquences de ce mécanisme, nous avons simulé ce groupe en déconnectant la possibilité d'une telle influence. La Figure 8.1a montre que cette influence a permis en effet à l'expert Taxon d'être sélectionné majoritairement lors des sessions avec plateforme visible (sessions 1 à 9), alors que, sans cette influence, l'expert Local aurait été majoritairement choisi (Figure 8.1 b). L'expert Taxon a donc pu prendre un plus grand contrôle sur le comportement le jour du test, avec plateforme visible déplacée (session 10).

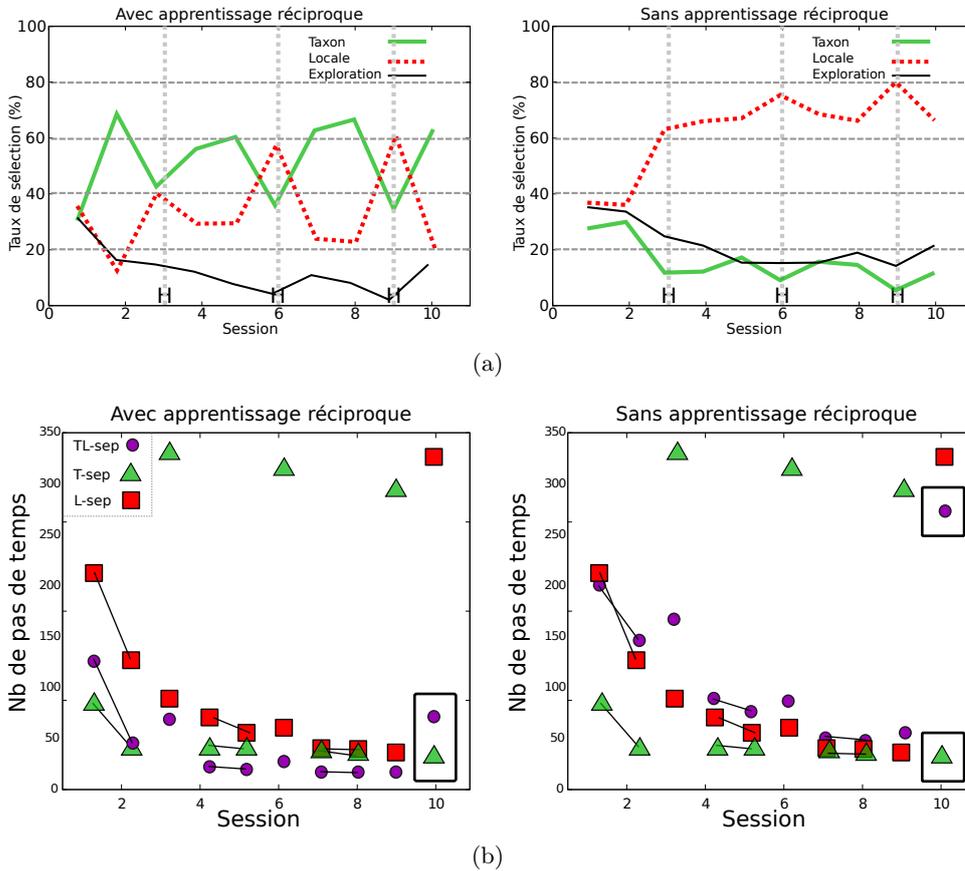


Figure 8.1 – Simulation de l’expérience de Devan et White (1999). Groupes TL-sep. (a) Résultats avec influence réciproque des apprentissages : le groupe Contrôle montre un temps d’échappement lors de l’essai de test semblable à celui des résultats observés. (b) Résultats sans influence réciproque : le temps d’échappement dans l’essai de test est non conforme, similaire à celui obtenu par le groupe P-sep.

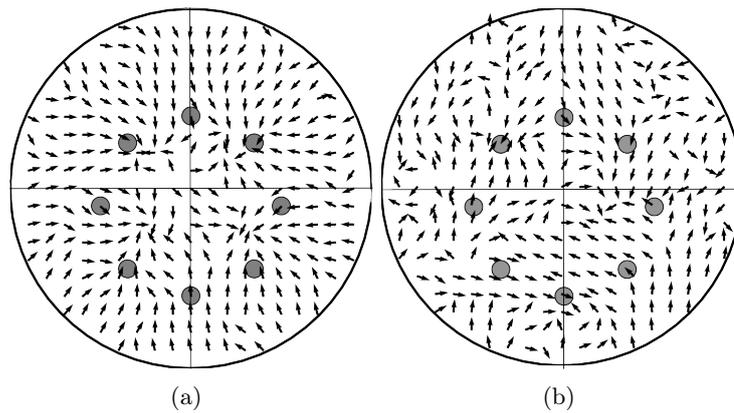


Figure 8.2 – Simulation de l'expérience de Roberts et Pearce (1999). Carte de navigation de l'expert Local dans le groupe DL-sep : (a) avec influence réciproque des apprentissages : les différents emplacements de la plateforme ont été mémorisés ; (b) sans influence réciproque : peu de lieux ont été mémorisés efficacement

Ce mécanisme favorise la coopération entre experts, par laquelle l'apprentissage des deux peut devenir plus performant qu'un seul. Nous pouvons par exemple démontrer que le groupe Contrôle DL-sep de l'expérience de Roberts et Pearce (1999) ("non conforme") n'avait pas pu produire de blocage observé des amers distaux, car la construction de la carte de navigation de l'expert Local, pourtant non sélectionné, avait été améliorée par l'expert Direction, contrôlant le comportement pour rejoindre un but visible déplacé. Les effets d'une telle influence ont pu s'estimer en déconnectant les effets comportementaux de l'expert Direction sur l'apprentissage de l'expert Local. Grâce à cette influence, la carte de ce dernier avait en effet pris en compte plusieurs emplacements possibles du but (Figure 8.2 a). Ainsi, le jour de test, lorsque le but ne pouvait être localisé de façon efficace qu'avec les indices distaux, l'expert Local avait pu contribuer à atteindre la plateforme. Avec un apprentissage réciproque déconnecté, sa carte de navigation n'aurait pas été construite de façon aussi élaborée à la fin de l'entraînement, et le blocage aurait pu apparaître (Figure 8.2 b).

Il est difficile toutefois d'évaluer précisément ces influences réciproques dans les explications des résultats obtenus, sans avoir au préalable comparé des groupes "connectés" ou "déconnectés". Toujours est-il que cet aspect du modèle s'avère important à conserver pour inciter des interactions coopératives.

8.2.3 Sélection associative adaptative

L'efficacité de la sélection est avant tout due au choix de la "monnaie commune" permettant au réseau de comparer l'efficacité de toutes stratégies, déterministes ou aléatoires. C'est grâce à cet atout, principale nouveauté de ce modèle, que l'on a pu prendre en compte ces diverses conditions expérimentales.

Pour éventuellement permettre la compétition entre modules, la sélection du réseau est apprise par un apprentissage par renforcement dont le fonctionnement est analogue à celui d'une règle d'apprentissage associatif (Sutton, 1988; Montague et al, 1995). Ce type d'apprentissage est relativement lent à apprendre et inflexible. Cette dernière propriété est responsable dans notre modèle des effets de blocage et d'occultation, mais aussi de gradient entre amers dans les expériences que nous avons simulées (Roberts et Pearce, 1999; Rodrigo et al, 2006).

En effet, le blocage des amers distaux par les proximaux dans l'expérience de Roberts et Pearce (1999) peut être expliqué par la valeur de sélection de l'expert de lieu qui va décroître durant 12 sessions, durant lesquelles le but est déplacé régulièrement, plus ou moins fortement selon que le but est déplacé tous les essais ou toutes les sessions. Lors des sessions suivantes durant lesquelles le but est stable, l'apprentissage du réseau de sélection est trop lent pour réévaluer la valeur de sélection de l'expert de lieu en seulement trois sessions – excepté dans les groupes Diff où le changement de balise dévalue la stratégie de guidage, augmentant mécaniquement les chances de l'expert de lieu d'être sélectionné.

De même, si l'apprentissage du réseau de sélection avait été plus rapide, l'effet de gradient d'occupation de l'octant de l'amer proximal provoqué par la distance avec l'amer distal dans l'expérience de Rodrigo et al (2006) n'aurait pas été obtenu. Ne trouvant pas la plateforme à l'endroit de l'amer proximal déplacé avec un expert de guidage (pendant un seul essai) , les animaux auraient immédiatement sélectionné l'expert de lieu et ce, quelle que soit la distance entre les deux amers.

De par l'apprentissage existant au niveau du réseau de sélection, prenant en compte directement les différentes entrées sensorielles qui sont fournies au réseau et aussi l'historique de la sélection (par le Q-learning qui y est appliqué), le réseau est très sensible aux conditions environnementales. On rappelle la simulation de l'expérience de (Devan et White, 1999), en laissant visible la plateforme dans les 3 sessions (sur 9) où elle était invisible, modification qui a conduit le groupe Contrôle à des performances bien supérieures qu'elles ne l'avaient été auparavant.

8.3 Contributions du modèle aux hypothèses relatives à l'apprentissage spatial

Le modèle peut donc rendre compte d'effets comportementaux relatifs aux trois conceptions de l'apprentissage spatial (associative, cognitive, modulaire). Nous verrons d'abord quels sont les apports généraux qui peuvent en être tirés, puis les apports particuliers spécifiques à chaque expérience simulée. Précisons à nouveau que ces contributions sont évidemment dépendantes du petit nombre d'expériences reproduit ici.

8.3.1 Apports généraux

Hypothèse d'amers indépendants en compétition. Tout d'abord, notons l'impossibilité des configurations du modèle ne supposant qu'une compétition d'amers (les groupes T-par ou D-par) à reproduire les expériences. La seule exception concerne un expert Direction, qui a pu reproduire très correctement *l'absence d'occultation* des amers proximaux par les distaux dans

l'expérience de McGregor et al (2004a) (Tableau 8.1, G), ce qui démontre que la prise en compte d'amers indépendants n'aboutit pas forcément à des effets associatifs. Nous reviendrons toutefois sur ce point dans une des sections suivantes.

L'hypothèse d'indépendance des amers vis-à-vis du contrôle du comportement peut être étayée par le fait que, parmi les amers proximaux, une balise (Exp1, Devan et White, 1999; Roberts et Pearce, 1999; Rodrigo et al, 2006, Tableau 8.1, H, F, E) acquiert toujours une valeur associative plus grande qu'un amer proximal situé à côté du but (Exp2, Pearce et al, 1998; McGregor et al, 2004a, Tableau 8.1, I, G). Cela est dû au fait qu'un expert sera d'autant plus choisi qu'il sera plus précis et, s'il pointe sur une balise, il proposera des orientations qui auront toute efficacité pour atteindre le but.

Dans toutes les simulations, les amers proximaux se sont révélés plus prégnants que les amers distaux. L'animat peut en effet rejoindre précisément le but dans toutes les directions autour de l'amer proximal, contrairement à ce qu'il peut faire vis-à-vis des amers distaux, ce qui limite la sélection de l'expert associé.

Ayant accès à l'amer proximal et l'amer distal dans l'expérience de Rodrigo et al (2006) (Tableau 8.1, E), un expert Direction a pu aussi produire, par compétition entre les deux amers, un gradient du taux d'occupation de l'octant de l'amer proximal déplacé. Dans l'expérience de Morris et al (1982) (Tableau 8.1, C), la recherche d'une plateforme cachée avec la seule aide des amers distaux a également pu être réalisée par un expert Direction par la compétition entre ces amers (s'ancrant seulement sur les plus petits), toutefois avec une performance moins bonne qu'un expert intégrant les amers (l'expert Local).

Si l'hypothèse de compétition entre amers indépendants ne suffit pas à rendre compte de tous les comportements, une totale intégration d'amers (e.g., groupes L-par ou P-par), ne suffit pas non plus à expliquer toutes les expériences (Tableau 8.1, colonne "non conformes").

Cette intégration, réalisée par les experts de lieu Local et Planification sur la base de cartes non-topologique ou topologique (surtout sur la base d'amers distaux), s'est cependant révélée *indispensable* pour rendre compte de certaines conditions expérimentales. Comparé à un expert de guidage qui aurait à sélectionner séquentiellement quelques amers pour atteindre le but, un expert utilisant les cellules de lieu intégrant ces quelques amers est en effet beaucoup plus rapide pour proposer une orientation pertinente.

Ce type d'expert a pu rendre compte, par un apprentissage rapide de la localisation de la plateforme par les seuls amers distaux, de l'amélioration importante de performance en peu d'essais (Pearce et al, 1998; Steele et Morris, 1999, Tableau 8.1, I, D) ou de la diminution de la saillance d'une balise, voulue par Rodrigo et al (2006) (Tableau 8.1, E). Il a pu également être rapidement sélectionné quand une plateforme visible pendant deux sessions a brusquement été cachée (Devan et White, 1999, Tableau 8.1, H). Dans l'expérience de Roberts et Pearce (1998) (Tableau 8.1, F), un groupe ayant accès à une balise et aux amers distaux (Session Same) a pu utiliser ces derniers dans une situation où l'amer proximal était absent grâce à la sélection d'un expert Planification. Un groupe similaire ayant utilisé seulement l'expert Direction ayant accès à tous les amers a en effet donné trop de valeur associative à la balise.

Hypothèse d'apprentissages différentiels. Rappelons que trois experts (Taxon, Direction, Local) ont un apprentissage procédural, un seul (Planification) a un apprentissage déclaratif. La majorité des simulations a nécessité le concours d'un expert apprenant sur un mode procédural, plus lent mais plus performant, après quelques essais, que le mode déclaratif. Un protocole programmant un temps d'entraînement important pendant lequel le but n'est pas déplacé avantage ce type d'expert. C'est pourquoi, paradoxalement, l'expérience classique de Morris et al (1982), étayant l'hypothèse d'une carte cognitive, a pu être mieux représentée par un expert procédural (Local), avec des performances meilleures qu'avec un expert déclaratif. Cela préserve cependant l'hypothèse de l'implication de structures nerveuses dépendantes de l'hippocampe (O'Keefe et Nadel, 1978), puisque l'expert Local s'appuie sur une carte de cellules de lieu. Toutefois la définition d'une carte cognitive pourrait être précisée quant aux notions qui y sont associées, comme la topologie et la flexibilité de son utilisation.

Peu d'expériences avec présence des deux types d'amers ont nécessité le seul concours d'experts procéduraux, sauf le comportement des rats Contrôles de Devan et White (1999), qui semble être mieux reproduit avec deux experts aux mêmes modes d'acquisition procéduraux (Tableau 8.1, H). Là encore, la plateforme est fixe pendant plusieurs sessions et favorise l'utilisation d'une stratégie plus automatique. D'autres comportements peuvent être bien représentés aussi avec ce seul type d'apprentissage (e.g., McGregor et al, 2004a; Rodrigo et al, 2006, Tableau 8.1, G, E), mais n'expliquent pas mieux les résultats qu'avec d'autres modes d'acquisition.

En revanche, la présence d'un expert Planification a été nécessaire lorsque le but est souvent déplacé, notamment dans l'expérience de Steele et Morris (1999) et de Pearce et al (1998) (groupe Contrôle) (Tableau 8.1, D, I), où un repositionnement rapide du but dans une représentation spatiale étaient indispensable. La tâche de Pearce et al (1998) a pu être reproduite sans expert déclaratif (groupe D), mais aucun expert procédural n'a pu améliorer de façon significative, comme l'expert Planification, la performance globale des animats conformément à celle des animaux. Pour un groupe de l'expérience de Roberts et Pearce (1999) également, qui n'a pas montré d'effet de blocage des amers distaux alors qu'il était entraîné sans ces amers (Groupe Trial-Diff), ce type d'apprentissage a aussi été indispensable pour qu'il assure le positionnement rapide de la plateforme en quelques sessions seulement (Tableau 8.1, F).

Ces remarques vont dans le sens de White (2004, 2005), insistant sur l'importance de prendre en compte l'existence des différents systèmes de mémoire pour comprendre le comportement spatial du rat.

Hypothèse de modules différenciés. La modélisation d'experts égocentrés et allocentrés, procéduraux et déclaratifs, a permis de retrouver un corpus de résultats variés sans avoir à modifier de façon importante les paramètres de ce modèle (voir aussi section 8.6). Cette robustesse étaye l'hypothèse de systèmes neuraux distincts, sous la dépendance de systèmes de mémoire différents, pouvant gérer diverses stratégies apprises en parallèle et pouvant être sélectionnées selon les circonstances. En grande majorité ce sont en effet les groupes composés de deux experts, plus spécifiquement avec entrées séparées, qui se sont révélés les plus fiables (Tableau 8.1, colonne "conformes").

Quelques exceptions cependant : la reproduction de protocoles où tous les amers contribuent

à l'atteinte du but et ne nécessitent pas l'utilisation de plusieurs modules (McGregor et al, 2004a; Devan et White, 1999, les jours où la plateforme est visible).

L'apparition d'effets associatifs entre amers (gradient, blocage et occultation) s'est révélée compatible avec l'hypothèse de plusieurs modules de navigation apprenant en parallèle. Elle peut être expliquée par les différences de performances entre les stratégies : l'utilisation d'une balise permet de toujours rejoindre un but sans avoir besoin de prendre en compte les autres éléments de l'environnement, tandis qu'un amer proximal placé à côté du but force l'animat à prendre en compte d'autres indices, ce qui se traduit dans le modèle par une sélection plus élevée de stratégies de lieu. Par exemple, l'obtention de blocage dans la simulation de l'expérience de Roberts et Pearce (1999) et l'absence d'occultation dans celle de McGregor et al (2004a) sont dus avant tout aux différences de contextes et de tâches effectuées : les premiers utilisent l'amer proximal comme balise, ce qui place d'emblée l'expert de guidage basé sur cet amer comme le plus pertinent pour résoudre la tâche, tandis que les seconds utilisent deux amers proximaux pour indiquer la plateforme, sans que celle-ci ne soit directement identifiable, où le concours des amers distaux s'avère utile pour localiser le but. Il a été déjà noté que la reproduction d'effets associatifs était très dépendante de la valeur prédictive des amers (e.g., voir l'expérience de Timberlake et al, 2007, qui montre qu'un tel blocage n'est possible qu'en utilisant l'amer proximal comme balise).

Nous avons montré qu'il est *aussi* possible qu'une coopération puisse émerger des interactions de tels modules (McDonald et White, 1993; Hartley et Burgess, 2005), y compris au cours d'une même expérience (e.g., Pearce et al, 1998). Dans ce protocole, la simulation de groupes n'exprimant qu'une des stratégies de lieu ou qu'une des stratégies de guidage s'est avérée moins performante que celle d'un groupe contrôle constitué des deux experts. Afin d'atteindre une plateforme invisible et déplacée pendant plusieurs sessions de 4 essais, ces deux stratégies étaient en effet nécessaires. A la fin de chaque session, en début d'expérience, la stratégie de guidage était compensée par la stratégie de lieu, l'expert de lieu apprenant beaucoup plus vite à relocaliser une plateforme déplacée. En fin d'expérience, la stratégie de lieu était compensée par la stratégie de guidage : ayant enfin appris à se diriger précisément vers l'amer proximal, l'expert de guidage pouvait conduire plus efficacement l'animat vers le but que l'expert de lieu, dont la carte de navigation était moins précise.

Même lorsque le module de sélection privilégie pendant plusieurs sessions l'une des stratégies, l'autre n'est jamais totalement négligée. C'est le principe fondamental du modèle, qui gouverne l'obtention d'effets associatifs, de compétition ou de coopération. Par exemple, dans l'expérience de Roberts et Pearce (1999), alors que les animats présentant un blocage des amers distaux n'utilisent effectivement plus l'expert de lieu pour se diriger vers le quadrant du but, cette stratégie est de nouveau sélectionnée pour rechercher la plateforme absente. Cette capacité est à rapprocher d'une variante de l'hypothèse associative "radicale", l'hypothèse du "comparateur" (*comparator hypothesis*) proposée par Miller et Matzel (1988), s'inspirant lui-même de l'hypothèse attentionnelle de Mackintosh (1975) pour expliquer de tels phénomènes, plutôt que de les expliquer par un déficit d'acquisition (Rescorla et Wagner, 1972).

Elle propose en effet que certains amers ou groupes d'amers peuvent acquérir une valeur associative plus forte que d'autres, mais n'entraînant pas forcément un blocage de l'apprentissage

de tous les autres amers et donc leur dévaluation. Ainsi, si les premiers ne sont plus pertinents, ceux qui ont acquis une valeur associative un peu moins forte peuvent être utilisés sans ré-apprentissage. Les effets de blocage et de gradient seraient issus de la sélection des stratégies et non de leur apprentissage. Cela permettrait notamment aux stratégies, grâce à leur apprentissage réciproque, de pouvoir s'améliorer et être sélectionnées lorsque nécessaire (notamment dans le cas d'un conflit sensoriel, e.g., Devan et White, 1999).

Typologie des modules Nous avons fait l'hypothèse que les amers proximaux étaient spécifiques des experts de guidage, et les amers distaux des experts de lieu. Il est avéré que les tâches dans lesquelles les deux types d'amers étaient présents, et n'étant pas toujours tous deux pertinents pour indiquer la position du but, ont pu être mieux exécutées par des experts à entrées séparées (Tableau 8.1).

Le traitement de ces amers – indépendants pour les amers proximaux, intégrés pour les amers distaux –, a été imposé a priori aux experts et a contribué pour une grande part à la complémentarité des stratégies utilisées. Cela rejoint la dichotomie souvent avancée entre systèmes neuraux indépendants ou dépendants de l'hippocampe, les seconds reposant sur l'utilisation d'une représentation par cellules de lieu (et graphe de planification, construit à partir de ces cellules) (e.g., Packard et McGaugh, 1992; White, 2004, 2005).

Contrairement à un usage très répandu (e.g., O'Keefe et Nadel, 1978; DeCoteau et Kesner, 2000; Killcross et Coutureau, 2003; Doeller et al, 2008), cette distinction entre systèmes *intégrant ou non les amers* (lieu et guidage) ne serait *pas superposable* à la dichotomie concernant les cadres de références ou les modes d'apprentissage.

En effet, la distinction entre stratégies de guidage et de lieu ont peu correspondu ici avec cadres *égocentré versus allocentré*. Au contraire, le choix d'experts allocentrés (Direction et Local ou Direction et Planification) a été beaucoup plus généralisé. L'expert Taxon égocentré se révèle en effet très performant pour positionner un but avec une balise, mais est rapidement dévalué lorsque le but est indicé autrement. Sa présence induit beaucoup plus de compétition entre experts qu'un expert Direction, car ses propositions d'orientation dans ce cas sont plus précises. Dans l'expérience de Roberts et Pearce (1999), dans laquelle la plateforme était pourtant balisée, un expert Taxon n'a pas pu rendre compte des résultats observés, provoquant soit un blocage total des amers distaux (TP-par ou T-par), soit une absence de blocage (TP-sep, Tableau 8.1 F). Dans le premier cas les amers distaux ancrés sur le Taxon ont tous été trop dévalués, dans le second cas la compétition entre experts Taxon correspondant aux deux amers proximaux possibles a trop favorisé la sélection de l'expert Planification. L'expert Direction a en revanche été utilisé dans beaucoup de protocoles, sa référence allocentrée lui accordant un avantage qui a permis, dans des situations où la plateforme était non balisée mais indicée, d'accorder aux amers proximaux une meilleure complémentarité avec les amers distaux.

La dichotomie entre stratégies de guidage et de lieu n'a pas non plus toujours correspondu avec apprentissage *procédural versus déclaratif*. Comme nous l'avons souligné plus haut, les situations où le but est à la même position durant l'entraînement ne nécessitent pas de la part de la stratégie de lieu une flexibilité particulière. La sélection de l'expert de lieu Local est donc favorisée car, après quelques sessions, il est plus performant que l'expert Planification. C'est le cas des

protocoles de Morris et al (1982) ou de Devan et White (1999) (pendant l'entraînement), dont les auteurs pourtant évoquaient l'utilisation de stratégies déclaratives.

Ces considérations peuvent apporter un éclairage différent sur la typologie des stratégies animales, qui souvent associe des cadres de référence ou des modes d'apprentissage spécifiques aux stratégies supposées reposer ou non sur l'intégration de certains types d'amers.

La distinction accordée par l'animal aux amers "à différencier" et d'autres "à intégrer" pourrait élargir à l'hypothèse de "combinaison adaptative" (*adaptive combination* Newcombe et Huttenlocher, 2006; Newcombe et Ratliff, 2007; Ratliff et Newcombe, 2008).

Fondée sur une vue associative de l'apprentissage spatial, cette hypothèse suggère que l'animal pondère a priori différents types d'indices selon un processus bayésien, avant d'être pris en compte dans la navigation. Ainsi, à la fois l'environnement, de par la nature des amers disponibles et l'historique de l'animal, vont influencer le pré-traitement de leurs entrées sensorielles. Apprenant au cours de leur vie (voire au cours de l'évolution, Ponticorvo et Miglino, 2010) à retenir ou à négliger certains types d'amers, car utiles ou trop peu utiles dans le passé, les animaux font ainsi par la suite l'économie de processus cognitifs coûteux pour se concentrer sur les éléments de l'environnement déjà connus dans le contexte.

Selon les auteurs proposant cette hypothèse, elle permettrait notamment d'expliquer les résultats discordants concernant l'utilisation de configuration géométrique par les animaux sans avoir à faire appel à un module géométrique "inné", mais plutôt à l'expérience de l'animal qui aurait appris que la géométrie était importante pour rejoindre un but. Les animaux n'ayant pas eu cette expérience se focaliseraient au contraire sur les amers usuels (amers proximaux ou distaux, Ratliff et Newcombe, 2008).

Précisons toutefois que les tenants de cette vue ont tendance à réfuter l'existence de modules spécialisés, tandis que nous y voyons au contraire l'opportunité de spécialiser certains modules en fonction des informations environnementales. Les animaux pourraient ainsi avoir appris quels amers seraient ou non à intégrer dans une représentation spatiale, indépendamment du mode de leur apprentissage et du cadre de référence des orientations choisies.

8.3.2 Apports plus spécifiques du modèle pour les protocoles reproduits

Des apports supplémentaires concernent plus particulièrement certains protocoles analysés ici, où pourraient être confirmées quelques prédictions du modèle.

Possibilité d'une représentation spatiale pauvre? Dans le chapitre 5, nous avons vu dans les deux premières expériences qu'il suffisait d'un seul amer pour avoir une représentation topologique utilisable, bien que moins complète que celles obtenues avec plusieurs amers. Nous fournissons en effet au modèle hippocampique des informations allothétiques pour la perception de l'amer, de même que les coordonnées de l'animat (informations idiothétiques, pour le calcul de l'activité des cellules de grilles), qui ne sont pas dépendantes de l'environnement. Malgré une certaine cohérence dans l'efficacité du modèle par rapport à la richesse de l'environnement (i.e., la carte construite en fonction des seules informations idiothétiques est de bien moindre

qualité que la carte construite en fonction d'un amer, elle-même moins détaillée qu'une carte construite en fonction d'une constellation d'amers), notre suggestion de la contribution d'une représentation spatiale en présence d'un environnement pauvre en amer serait à approfondir.

Dans l'expérience EXP2, dans laquelle une plateforme était indiquée par un seul amer, l'utilisation d'une telle carte a beaucoup amélioré la performance des groupes TP-par et DP-par (Tableau 8.1 B), jouant très certainement un rôle exploratoire. Nous suggérons que l'inactivation de l'hippocampe chez des animaux (censés ainsi ne plus pouvoir construire de représentation spatiale) pourrait alors aboutir à des performances moins bonnes qu'un groupe Contrôle.

On a vu aussi que ce type de carte, construite sur un seul amer distal, contribuait spécifiquement aux tests d'extinction dans l'expérience de Rodrigo et al (2006), et donc à la diminution de l'occultation de cet amer voulue par les auteurs. Là encore, l'absence d'utilisation d'une carte hippocampique pourrait empêcher l'occupation de l'octant de l'amer distal pendant ces tests.

Par ailleurs, dans cette même expérience, dans les essais tests où l'amer proximal est déplacé, le modèle prédit que le gradient d'occupation de cet octant ne devrait pas s'obtenir jusqu'à la première atteinte de cet octant. L'effet de gradient ne devrait s'établir qu'*ensuite*, grâce à l'utilisation de cette carte cognitive, les animaux se dirigeant alors vers l'octant de l'amer distal. Les auteurs n'ayant pas distingué ces deux étapes dans l'obtention d'un gradient, il serait possible de le tester dans de futures expériences.

Contribution des amers distaux dans l'expérience de Roberts et Pearce (1999), malgré un blocage de ces amers. Ces deux étapes dans l'obtention d'un effet associatif se retrouvent aussi dans cette expérience. L'effet de blocage des amers distaux, dans le groupe "Trial-Same" s'étant entraîné avec la même balise et ayant été, pendant quelques essais seulement, en présence d'amers distaux, *ne s'exprimerait que* jusqu'à l'atteinte, par hasard, de la zone de la plateforme absente. Ensuite, ces amers seraient de nouveau utilisés pour rechercher cette plateforme. Un protocole spécifique pourrait être conçu pour étayer cette hypothèse.

Possibilité d'obtention d'un effet d'occultation dans l'expérience de McGregor et al (2004a). Dans le protocole de cette expérience, tout était établi (avec une plateforme fixe et non balisée en présence d'amers proximaux et distaux) pour qu'une absence d'occultation des amers proximaux par les distaux ait lieu dans le groupe Stable. En effet, les stratégies de guidage et de lieu se sont avérées redondantes. Les auteurs avaient supposé qu'en continuant l'entraînement, ils auraient obtenus une telle occultation.

La seule utilisation d'une stratégie Direction (groupe D-par) avait permis d'obtenir une absence d'occultation et nous avons indiqué, à propos de ce groupe, qu'étant donné l'augmentation de la saillance des amers distaux et la dévaluation de l'un des amers proximaux, il était possible en effet qu'un entraînement plus long ait cet effet. Nous avons donc simulé un groupe D-par Stable en lui accordant, non pas 12 sessions d'entraînement, mais 20. La Figure 8.3 a montre que la sélection des amers distaux atteint un palier après environ 10 sessions et n'augmente plus, alors que la sélection des amers proximaux diminue progressivement, au profit d'une augmentation de l'exploration. Les animats ont donc capitalisés petit à petit sur ce comportement aléatoire et sur les amers distaux plutôt que de positionner la plateforme par les amers proximaux. Les tests en position "opposée" de la plateforme semblent avoir provoqué cette dévaluation des amers proximaux au profit de l'exploration.

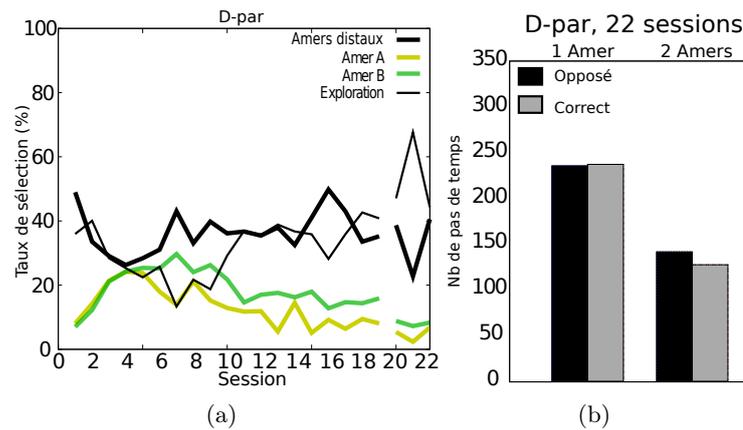


Figure 8.3 – Simulation de l'expérience de McGregor et al (2004a) avec entraînement prolongé sur 20 sessions pour le groupe D-par. (a) Taux de sélection des experts au cours des sessions (b) Résultats aux essais de test.

Aux sessions test 1- ou 2-amers, les animaux présentent en effet une différence non significative des temps d'échappement pour rejoindre la plateforme avec les seuls amers proximaux, placée à l'endroit correct et opposé (Figure 8.3 b). On peut donc en déduire qu'un entraînement beaucoup plus long aurait en effet conduit, comme supposé par les auteurs, à une occultation des amers proximaux, malgré un protocole qui n'était pas propice à une telle démonstration. Cependant la différence significative des temps d'échappement entre les situations 1- et 2-amers indique que l'occultation de ces deux amers n'est pas encore établie très fortement.

Influence de l'historique de la sélection dans l'expérience de Devan et White (1999).

Les animaux avaient à retrouver dans cette expérience une plateforme visible déplacée après un entraînement de 6 jours pendant lesquels la plateforme était visible et fixe, intercalé avec 3 jours où elle était invisible. Le jour du test de compétition où la plateforme est déplacée, le groupe Contrôle avait une performance intermédiaire entre un groupe Taxon (meilleur) et un groupe Local (moins performant). La présence de seulement 3 jours de sessions pendant lesquelles la plateforme était invisible a contribué de façon importante à l'obtention des résultats intermédiaires du groupe Contrôle. Nous avons montré en effet que, sans ces sessions, ce groupe aurait eu des performances meilleures, similaires au groupe Taxon (Figure 7.7). L'historique des sélections démontre donc à quel point ces sessions ont empêché l'expert Taxon de prendre un contrôle trop important. Un protocole spécifique pourrait vérifier ces résultats chez les animaux.

Apport de la simulation d'un groupe "lésé DLS" dans l'expérience de Pearce et al (1998).

Dans un groupe Contrôle devant rejoindre une plateforme indiquée et déplacée tous les 4 essais, nous avons démontré une compétition entre stratégie de guidage et de lieu aux premiers essais de chaque session et une coopération entre ces stratégies aux quatrièmes essais. La première hypothèse avait été déjà suggérée par les auteurs, mais pas la seconde. Nous avons pu

étayer nos déductions par la simulation d'un groupe "lésé DLS", supposé ne pas développer de stratégie Direction, mais n'ayant pas été prévu dans l'expérience. En effet, ce groupe montre clairement une absence d'apprentissage inter-sessions et un apprentissage très rapide intra-session. La comparaison de ces résultats avec un groupe "lésé Hc", dont les résultats étaient conformes à ceux testés dans l'expérience réelle, a pu conduire à ces hypothèses d'interactions antagonistes de stratégies au cours d'une même expérience. Le test d'un tel groupe de rats lésés DLS pourrait conforter ces hypothèses.

8.4 Eléments limitants du modèle

Comme nous l'avons vu, une grande partie des résultats obtenus l'ont été sur la base du critère de sélection. Or ces résultats dépendent de biais qui peuvent provenir d'une simplification des entrées sensorielles des stratégies, de la nature même de leur apprentissage ou du mécanisme de sélection.

8.4.1 Traitement sensoriel des amers

Les "cellules sensorielles" dont nous faisons état et qui sont utilisées pour construire les cellules de lieu sont en fait des cellules percevant déjà un amer comme étant distinct des autres objets de l'environnement. Or, ces amers font l'objet d'un traitement sophistiqué par les animaux avant d'être effectivement utilisés pour la navigation. Cheng et al (2007) fait une revue d'un certain nombre d'expériences dans lesquelles une pondération bayésienne (i.e., les amers de moindre variance ont une saillance plus importante) pourrait expliquer pourquoi, dans certains cas, les amers sembleraient être intégrés et, dans d'autres, en compétition. Ainsi, comme nous l'avons déjà évoqué en mentionnant l'hypothèse comparative, la distinction habituelle entre les amers distaux et proximaux pourrait être prise en compte dans une intégration bayésienne de ces amers, en amont de leur "récupération" par les stratégies de navigation. Cela impliquerait ainsi que la sélection des stratégies sur la base de leur seule performance ne peut pas expliquer par exemple pourquoi un amer est préféré à un autre, et donc qu'une telle préférence pourrait être pondérée en amont.

L'influence directe d'une telle simplification sur la sélection n'est cependant pas évidente à évaluer, elle peut surtout rendre certaines stratégies plus faciles (ou plus difficiles) à apprendre, rajoutant une hypothèse sur leur apprentissage qu'il n'est pas aisé de vérifier expérimentalement.

8.4.2 Mécanismes d'apprentissage simplifiés

Stratégies de guidage. Des biais concernant la modélisation de ces stratégies peuvent être dus aux conditions idéales de la simulation. Nous avons mentionné que l'apprentissage de la stratégie de guidage provoque des réponses d'autant plus précises que l'amer est petit. Cela est dû notamment au fait que la relation entre la taille de l'amer et du renforcement des actions est linéaire (Figure 5.13) : un amer plus petit permettra une navigation plus précise. Or, ce n'est pas forcément le cas chez les animaux, qui ont au contraire tendance à retenir les amers de taille plus importante, ces derniers étant reconnus comme plus stables (e.g., Gouteux et al, 2001). Si,

pour les amers distaux simulés, notamment de par leur position pseudo-aléatoire, l'influence d'un tel biais est négligeable, il est important de noter que dans le cas de l'expérience de McGregor et al (2004a) où les deux amers proximaux ont des tailles différentes, un tel biais aurait pu être dommageable si les amers n'étaient pas alternés selon les expériences. Toutefois, ce biais n'est pas complètement déraisonnable, compte tenu de l'observation de Cheng et al (2007), notée dans le Chapitre 3, qu'un groupe d'amers plus petits peut être préféré à un amer de plus grande taille, du fait que ces premiers pris ensembles peuvent être plus stables.

Par ailleurs, les stratégies Taxon et Direction ne peuvent utiliser plusieurs amers que séquentiellement, au travers d'un réseau de sélection dédié. C'est ainsi que l'expert Direction, dans l'expérience Morris et al (1982), a pu résoudre la tâche (comme les animaux Hc), sans pour autant intégrer les amers distaux, en inférant une direction générale à partir de plusieurs amers proches. Pour des raisons pratiques, nous n'avons pas modélisé de stratégie de guidage pouvant apprendre en amont du réseau de sélection une orientation en fonction de plusieurs amers. Par exemple, nous pourrions sommer les encodages d'amers distincts et réaliser l'apprentissage à partir des cellules visuelles ainsi créées. La Figure 8.4 (a) montre la structure actuelle des experts Direction ou Taxon pour plusieurs amers : une présélection est effectuée entre tous les experts Direction ou Taxon par un réseau de sélection intermédiaire. Ce réseau et les experts représentent l'expert Direction ou Taxon global. Celui-ci est ensuite sélectionné parmi les autres experts (lieu, exploration). En revanche, la Figure 8.4 (b) montre un expert Direction ou Taxon où les amers présents sont directement reliés aux actions d'orientation par des poids synaptiques sur lesquels le Q-learning sera effectué. Cette dernière structure n'apprend ainsi qu'une seule orientation par pas de temps pour les deux amers présents, tandis que la première en apprend deux, pour chaque amer. La modélisation d'un tel Taxon se rapprocherait alors de la stratégie de guidance précédemment mentionnée (Chapitre 2, section 2.2.2)

Ce mécanisme pourrait améliorer la modélisation de ces stratégies, cependant il aurait fallu pour cela déterminer le nombre maximum d'amers pouvant être appris par un seul expert, mais également considérer les facteurs de "regroupements" de ces amers (e.g., par distance, ressemblance, etc.) et tester toutes les combinaisons possibles d'experts de guidage pour en déterminer les plus probables, ce qui, dans le cas d'une dizaine d'amers, se serait révélé impossible.

Stratégies de lieu. La modélisation des stratégies de lieu a été, pour les besoins de la simulation, grandement simplifiée. Nous avons mentionné précédemment que les cartes des experts de lieu sont toujours créées avant le début de l'expérience, et ne sont pas réappries ensuite. Cela suppose l'existence d'une période d'exposition préalable de l'animal dans l'environnement de test. Ces périodes sont parfois effectivement réalisées (e.g., Rodrigo et al, 2006, en l'absence d'amers) ou non (e.g., Roberts et Pearce, 1999). Nous faisons donc ici l'hypothèse que la construction d'une telle représentation s'avère très rapide et peut être opérationnelle en quelques essais. Nous considérons cette hypothèse raisonnable, compte tenu de l'exhibition rapide par les rats de telles capacités (e.g., O'Keefe et Nadel, 1978).

Nous ne pouvons cependant toutefois pas statuer, comme par exemple Avni et al (2006), sur le moment de la construction d'une telle représentation spatiale. Ils suggèrent en effet que les animaux utilisent au fur et à mesure de leur présence dans un environnement nouveau des stratégies de complexité incrémentale : d'abord un taxon simple pour se rapprocher des amers

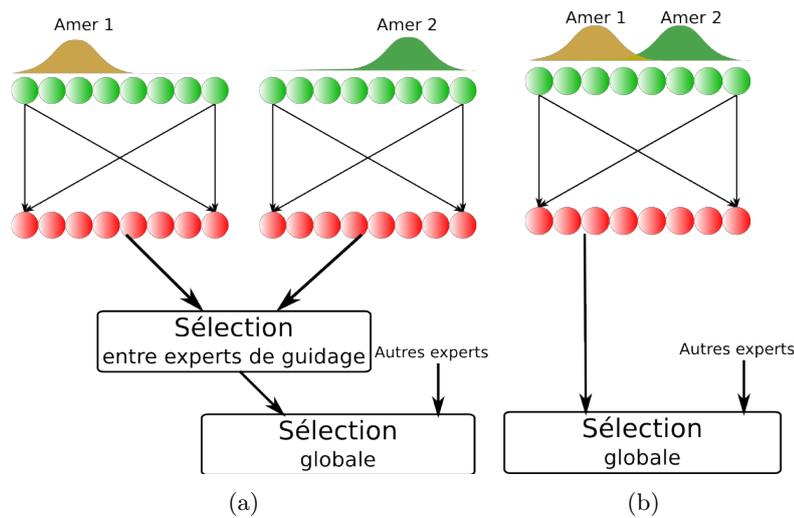


Figure 8.4 – Différentes possibilités d’implémenter une stratégie de guidage avec deux amers : (a) celle que nous avons choisie pour le modèle consiste à attribuer un expert pour chaque amer, et à ensuite opérer une sélection sur les orientations proposées par chaque expert ; (b) on pourrait aussi sommer les encodages des amers et y appliquer directement l’apprentissage.

les plus saillants, puis une direction à partir d’un amer, enfin une représentation spatiale (voir aussi Lehmann et al, 2007). Toutes nos simulations démontrent en revanche que l’utilisation d’une telle hiérarchie est très contextuelle, dépendant étroitement de l’environnement et des expériences passées.

Ces représentations spatiales ont été apprises par le modèle en fonction des amers distaux ou de tous les amers. Nous n’avons cependant pas pris en compte la possibilité de l’existence de deux types de représentations spatiales : l’une fondée sur les amers distaux et l’autre, sur les amers proximaux, comme semblent le suggérer l’existence de cellules de lieu sensibles à un seul type d’amer (Renaudineau et al, 2007). Il serait donc possible d’avoir deux types de stratégies de lieu différenciées par la nature des amers pris en compte. Une telle possibilité pourrait par exemple améliorer les simulations du groupe "lésé DLS" de l’expérience de Devan et White (1999), dont le comportement a été mal reproduit lors de l’entraînement. Nous avons en effet indiqué, dans la discussion partielle (section 7.1.3), que les animaux auraient pu utiliser deux sortes de stratégies, une Locale lors des sessions à plateforme visible, et une Planification lors des sessions à plateforme cachée. On pourrait supposer que ces deux stratégies de lieu, l’une utilisant des cellules de lieu construites avec les amers proximaux, l’autre des cellules construites avec des amers distaux, aient pu s’exprimer. Cela supposerait, dans ce cas précis, qu’elles n’auraient pas le même mode d’apprentissage.

8.4.3 Contraintes dans le réseau de sélection

Normalisation des entrées. Les entrées sensorielles des experts sont normalisées avant d'être transmises au réseau de sélection. Nous avons choisi ici une normalisation telle que la moyenne des activations des entrées d'un expert soit égale à celle des entrées des autres experts. Ainsi, l'importance d'un type d'entrée dans le processus de sélection ne devrait être attribuée qu'à sa pertinence effective pour la sélection et non un déséquilibre préalable de ses activations.

Malgré cette précaution, il apparaît pourtant que certains choix dans les entrées peuvent paraître surprenants : par exemple dans l'expérience de Roberts et Pearce (1999), le jour de test, alors que la balise indiquant le but est absente, c'est l'expert Direction dépendant de cette balise qui est choisi. Cela est dû notamment au choix de remplacer la vision de l'amer par du bruit lorsque ce dernier est absent (section 4.1.1.1). Or, ce bruit, une fois normalisé, s'avère aussi influent que les autres activations des entrées sensorielles.

Cette sélection peut trouver une explication biologique dans le fait que l'animat, ayant appris tout le long de l'expérience à se référer à cet amer, va continuer à le chercher pendant un certain temps après sa disparition, ce qui expliquerait aussi la possibilité de blocage.

Persistance de la sélection. Le réseau de sélection déterministe ne permet pas une sélection optimale des stratégies à cause d'un manque de persistance. De fait, il a été nécessaire d'imposer à notre réseau de sélection de garder durant trois pas de temps l'exploration lorsque cet expert venait à être choisi, et ce afin de lui permettre d'éloigner efficacement l'animat de son emplacement précédent. Des oscillations comportementales ont également affecté les experts, comme par exemple dans l'expérience de Pearce et al (1998) (trajectoire de la Figure 7.13 d), où stratégies Planification et Direction sont sélectionnées alternativement à proximité du but de façon peu optimale.

La persistance d'un comportement est l'une des problématiques investiguée par les modèles computationnels de sélection de l'action, notamment ceux inspirés par le fonctionnement des boucles cortico-basales (Gurney et al, 2001a; Girard et al, 2005, 2008). L'adjonction d'un tel modèle au nôtre, au niveau du réseau de sélection, pourrait permettre d'améliorer cette persistance, pour éviter par exemple de forcer la sélection de l'exploration, mais aussi de mieux contrôler la sélection de sorte à ce que le calcul inhérent à la sélection n'intervienne que lorsque nécessaire. Cette amélioration est confortée par l'intégration d'un tel modèle, qui est en cours. Les premiers résultats sont consignés dans le travail de Liénard (2009).

Choix a priori des experts et passage à l'échelle. Dans chacune des simulations, en plus de l'expert exploration, nous avons choisi un expert de lieu et un expert de guidage que nous jugions plus pertinents que d'autres mais, alors que c'était possible, nous n'avons jamais mis en présence deux experts de guidage ou deux experts de lieu. Ce choix a été imposé, d'une part, pour réduire le nombre d'analyses et, d'autre part, pour suivre la dichotomie classique dans les expériences de navigation (guidage versus lieu).

Il aurait sans doute été plus fiable de proposer au module de sélection tous les experts possibles et de laisser émerger les stratégies vraiment efficaces. Cependant la question de l'architecture d'une telle sélection se pose alors : faut-il sélectionner les modules "à plat" – tous les experts

étant sélectionnés au sein d'un seul réseau de sélection – ou au contraire induire une hiérarchie ? Quelle forme aurait cette hiérarchie pour qu'elle soit la plus souple ? Les stratégies procédurales doivent-elles d'abord être sélectionnées à part, puis l'orientation finale sélectionnée avec l'expert Planification ? Les sélections doivent-elles se faire au niveau du cadre de référence, ou du type d'entrées (lieu vs entrées sensorielles) ? Le modèle pourrait s'inspirer du principe de sélections en cascades sous-tendant le travail d'Haruno et Kawato (2006), qui proposent un modèle computationnel des ganglions de la base – non encore appliqué à la navigation et dévolus à des systèmes d'apprentissage par renforcement partageant le même type d'information – qui génère une hiérarchie de stratégies reposant sur des informations d'abord éparses puis plus détaillées pour réaliser une tâche de pointage.

En bref, le choix de telles architectures est encore à investiguer.

8.5 Assignation de structures nerveuses aux experts et au module de sélection

Comme on l'a mentionné dans notre objectif et illustré dans les Chapitres précédents, ce modèle bio-inspiré se place essentiellement au niveau comportemental pour expliquer, par les interactions de modules spécialisés, les effets de l'apprentissage spatial. Les équivalents des experts au niveau des structures nerveuses ont été repris à partir du Chapitre concernant les données biologiques, d'où nous avons tiré que les stratégies de lieu et de guidage seraient dépendantes de structures nerveuses différentes, impliquant ou n'impliquant pas la construction de cellules de lieu par l'hippocampe ; qu'elles pourraient avoir des cadres de références différents ; que leurs apprentissages différentiels pourraient émerger aux systèmes de mémoire procédural et déclaratif ; qu'ils seraient plutôt focalisés sur des types d'amers spécifiques ; que le réseau de sélection serait sous-tendu par une structure séparée de celles que sont responsables de l'apprentissage des stratégies.

Ces équivalents ont pu correctement reproduire des comportements d'animaux lésés, comme les experts Taxon et Direction pour les groupes d'animaux hippocampiques (Morris et al, 1982; Pearce et al, 1998; Devan et White, 1999), les experts Local ou Planification pour les groupes d'animaux lésés du DLS (Devan et White, 1999). Leurs sélections conjointes ont aussi pu correctement reproduire le comportement de rats intacts (Morris et al, 1982; Pearce et al, 1998; Devan et White, 1999; Roberts et Pearce, 1999; McGregor et al, 2004a; Rodrigo et al, 2006).

Cependant ces équivalences ne reprennent que les principes très généraux du fonctionnement de ces structures et certaines stratégies ne correspondent pas précisément aux données les concernant.

Experts de guidage. Si le DLS, avec le colliculus supérieur, semblent bien être impliqués dans l'apprentissage de la stratégie Taxon telle que nous l'avons reproduite (égocentrée et procédurale) (Packard et McGaugh, 1996; Devan et White, 1999; Goodale et al, 1978), la référence allocentrée de l'expert Direction impliquerait une autre configuration. Cette stratégie directionnelle, distincte d'une stratégie de lieu, a permis au modèle de retrouver une majorité de comportements qui auraient été sans doute attribués à une stratégie égocentrée. Elle est de plus

en plus souvent évoquée pour être utilisée fréquemment par les animaux (e.g., Pearce et al, 1998; Hamilton et al, 2007, 2008). Les cellules de direction de la tête, impliquant le noyau anterodorsal sont supposées fournir à la stratégie ce référentiel (Taube et al, 1990; Skinner et al, 2003). Il est toutefois difficile de préciser plus spécifiquement quels circuits nerveux y seraient impliqués.

Experts de lieu. L'apprentissage des stratégies de lieu est sous la dépendance de l'hippocampe (e.g., Morris et al, 1982; Packard et McGaugh, 1996) et probablement aussi du DMS (Devan et White, 1999). Le mPFC permettrait la flexibilité comportementale inhérente à la planification : Granon et Poucet (1995) ont montré en effet que la lésion de cette structure, si elle n'empêchait pas la stratégie de lieu d'être apprise, lui enlevait toute flexibilité.

Nous avons modélisé deux types de stratégies de lieu, reposant toutes deux sur une carte de cellules de lieu supposées être construites dans les régions hippocampiques et parahippocampiques, mais dont l'apprentissage diffère : la stratégie Locale exécute directement sur les cellules de lieu un apprentissage procédural, de sorte que cette stratégie est plus précise, mais moins flexible que la stratégie de Planification, fondée sur une carte topologique construite comme une représentation éparsée des cellules de lieux, permettant ainsi la création de lien entre différents lieux (Gaussier et al, 2007). Ainsi cette dernière stratégie est plus rapide pour localiser un but et planifier son chemin, de même qu'elle peut rapidement relocaliser ce but si celui-ci est déplacé, mais est plus imprécise que la stratégie Locale, de par la nature éparsée de sa représentation.

Cette dichotomie entre ces deux stratégies de lieu est très commune en navigation computationnelle (e.g., Trullier et al, 1997; Filliat et Meyer, 2003; Girard et al, 2005). Un modèle computationnel de navigation en a démontré les performances respectives (e.g., Busch et al, 2007). La distinction entre ces deux stratégies s'est avérée très utile pour certaines de nos démonstrations (e.g., Morris et al, 1982; Pearce et al, 1998; Steele et Morris, 1999; Devan et White, 1999), mais l'implication de structures nerveuses dédiées à une stratégie Locale reste cependant hypothétique.

Exploration. Bien que notre module consistant à choisir régulièrement une orientation aléatoire soit très peu plausible pour simuler un comportement exploratoire, la nature modulaire, et donc l'existence d'un système neural dédié d'un tel expert, est supportée par une revue récente (Chakravarthy et al, In Press), mais encore non attribuée précisément à une structure nerveuse. Des candidats potentiels seraient le locus cœruleus (Doya et al, 2002) ou des circuits entre le globus pallidus et le STN (Sridharan et al, 2006), systèmes modulés par la noradrénaline, neuromédiateur dont l'activité semble liée à l'exploration.

Module de sélection. Dans le réseau de sélection, afin de comparer les différentes orientations proposées par les stratégies, nous les avons toutes converties par commodité dans un référentiel allocentrique (le simulateur ayant accès aux données dans ce référentiel). Il est cependant plus probable, compte tenu de la nécessité pour les orientations d'être ensuite converties en ordres moteurs, que cette conversion s'effectue dans un référentiel égocentrique. Le mPFC (Kesner et al, 1989) et le DLS (Brasted et al, 1997; De Leonibus et al, 2005) contribueraient à la gestion égocentrique de la navigation. La coexistence des deux référentiels serait cependant possible par la présence du cortex pariétal associatif, qui contribuerait à la gestion d'un cadre allocentrique.

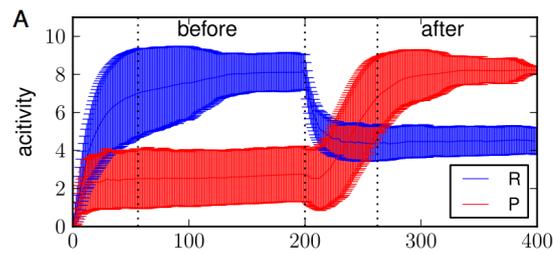


Figure 8.5 – Simulation de l’activation des populations de neurones correspondant aux unités de la stratégie réponse (R en bleu, procédurale comme une stratégie de guidage) et de lieu (P en rouge, équivalent du Local) lors d’un changement de tâche : les rats simulés doivent d’abord toujours effectuer le même virage pour rejoindre un bras d’un labyrinthe en croix (Before, stratégie réponse), puis trouver le but selon un bras indiqué par les amers distaux (After, stratégie de lieu). L’activation des unités suit le changement de tâche. Figure tirée de Sheynikhovich et al (2010).

Save et Moghaddam (1996) et Save (2000) proposent en effet l’hypothèse que le cortex pariétal associatif gère l’intégration multi-sensorielle et transforme ainsi l’information égocentrique en information allocentrique, en amont de l’hippocampe.

Comme il avait déjà été évoqué dans le Chapitre 6, le module de sélection pourrait correspondre aux zones IL-PL du mPFC, mises en évidence par, e.g., Ragozzino et al (1999) ou Rich et Shapiro (2009). Ces zones en effet ne seraient pas responsables de l’apprentissage des stratégies, mais de leur sélection. En outre, elles reçoivent des connexions afférentes de l’hippocampe (e.g., Conde et al, 1995) et du DLS (e.g., Groenewegen et al, 1991). Le mPFC reçoit de plus des projections dopaminergiques de l’aire tegmentale ventrale, fournissant la base nécessaire pour un apprentissage lié à une récompense, tel le Q-learning du module de sélection.

Rich et Shapiro (2009) ont relevé que différentes sous-populations de neurones du mPFC pouvaient coder les choix de stratégies comportementales différentes. Dans notre modèle, les valeurs associées à chaque stratégie par le module de sélection pourraient être considérées comme une représentation de l’activité de ces sous-populations. En effet, les changements entre stratégies correspondent à des changements relatifs de ces valeurs : si la valeur de la stratégie de guidage est plus grande que la valeur de la stratégie de lieu, alors la stratégie de guidage prendra le contrôle du comportement, et réciproquement (e.g., Chapitre 7). Les changements entre ces valeurs pourraient donc correspondre aux changements relevés entre les activités de populations de neurones dans les données enregistrées par (Rich et Shapiro, 2009, Figure 6a de leur article). Une version très simplifiée de notre modèle reproduisant un changement d’une stratégie de lieu vers une stratégie de guidage a testé cette hypothèse dans un labyrinthe en croix (Sheynikhovich et al, 2010). Les activations des unités du réseau de sélection montrent en effet des similitudes avec celles des populations de neurones des zones IL-PL (Figure 8.5).

Malgré ces similarités, le module de sélection ne peut pas correspondre à quelques autres données

relatives au rôle du mPFC dans la sélection des stratégies. Par exemple, il a été montré que les zones IL-PL sont responsables de sélections inter-modales, mais pas de sélections intra-modales (comme, par exemple, le *reversal learning* Young et Shapiro, 2009), qui serait sous la dépendance de l'OFC. Dans notre modèle cependant, ces deux types de sélections peuvent être apprises par le même réseau, puisque le reversal learning correspond simplement à un changement du lieu de la récompense. Une autre incompatibilité vient du travail de Rich et Shapiro (2007), qui ont montré que les zones IL-PL étaient impliquées seulement pendant les tout premiers changements de stratégies, mais pas dans les changements ultérieurs. Notre modèle ne peut pas fournir l'explication plausible de ces données.

En résumé, le mPFC pourrait être considéré comme une structure biologique plausible pour le réseau de sélection mais, dans ce cas : (i) une séparation du réseau en au moins deux parties est indispensable pour prendre en compte les données de sélections intra-modales et extra-modales et (ii) une extension du modèle est requise pour expliquer comment les stratégies intervenant après quelques changements sont sélectionnées.

8.6 Bilan : le modèle comme outil d'analyse et de prédiction

Malgré les limitations que présente ce modèle, tant du point de vue de la modélisation des stratégies de navigation qu'au niveau de leur mécanisme de sélection, il peut apporter des informations non négligeables sur la nature de l'apprentissage spatial, qui ne peuvent être simplement accessibles, même par des manipulations comportementales ou des lésions.

Identification des entrées sensorielles. Pour chaque expert, les entrées sensorielles sont connues et identifiables à n'importe quel moment de l'expérience ou d'une trajectoire individuelle. On peut ainsi estimer sur quel amer (ou quels amers) l'animat s'est ancré pour rejoindre un but (comme, par exemple, l'expert Direction dans l'expérience de Morris et al, 1982, Figure 5.14). Ces observations pourraient être vérifiées dans des expériences réelles par des manipulations comportementales ou des analyses de trajectoires, par exemple avec les méthodes d'Hamilton et al (2004) ou Harvey et al (2009), permettant d'identifier l'amer ou le type d'amer choisi par l'animal durant sa trajectoire en analysant les orientations de sa tête.

Caractéristiques des experts. La modélisation des experts selon un schéma précis (informations sensorielles, cadre de référence, nature de l'apprentissage) permet également d'identifier précisément quelles sont les stratégies les plus à même de remplir une tâche simulée en fonction de ce schéma et, en retour, d'apporter des hypothèses, quelquefois contre-intuitives, sur le type de stratégies utilisées par l'animal. Par exemple, malgré la présence d'une balise, une stratégie Direction a été préférée par les animats dans une tâche de blocage (Roberts et Pearce, 1999; Rodrigo et al, 2006).

Des comportements de groupes d'animaux lésés de structures neurales supposées être impliqués dans certains apprentissages (e.g., hippocampe pour la stratégie de lieu, DLS pour la stratégie de guidage) ont pu être reproduits en inactivant l'expert associé : celui-ci ne peut donc pas être sélectionné parmi les autres experts et n'influence pas l'apprentissage de l'autre. Il est

donc possible de proposer des lésions non testées par les expérimentateurs (par exemple, la lésion du DLS dans l'expérience de Pearce et al, 1998) et de démontrer également que les conséquences supposées de certaines lésions ne sont pas forcément aussi simples que ne le prévoit le modèle (cf. Devan et White, 1999).

Enfin, l'accès aux variables internes des experts donne lieu à la constitution de cartes de navigation pour chacun d'entre eux. Par exemple, nous pouvons montrer les différences dans les cartes qu'un expert Taxon et qu'un expert Direction peuvent construire par rapport aux deux amers proximaux dans l'expérience de McGregor et al (2004a) (cf Chapitre 7), ainsi que les différences entre les graphes de planification construits sur peu (Rodrigo et al, 2006) ou plusieurs amers (Pearce et al, 1998). De même, nous montrons comment des cartes de navigation peuvent être influencées par l'apprentissage parallèle d'un autre expert (cf. plus haut expérience de Roberts et Pearce, 1999).

Taux et chronologie de la sélection des experts. Les taux de sélection d'experts, dont on peut donner l'évolution par sessions ou au cours d'une trajectoire, permettent notamment d'estimer à quels moments les animaux sont susceptibles de changer de stratégie. Ils peuvent également démontrer comment l'historique des sélections influence le comportement des animaux en cas de conflits d'amers (e.g., Devan et White, 1999). Il est également possible de connaître la sélection des experts dans un lieu particulier. Nous avons pu ainsi déterminer, par exemple dans l'expérience de Rodrigo et al (2006), quels sont les experts les plus sélectionnés selon l'emplacement de l'animat par rapport à l'octant où se trouve l'amer proximal (Figure 6.3 e).

Influence des amers environnementaux sur la sélection d'experts. Il a été le plus souvent facile d'inférer l'influence des amers par les seuls taux de sélection d'experts à entrées séparées. Cependant, dans certains cas, cette inférence est difficile à effectuer. Le modèle peut alors fournir des données d'analyses complémentaires par les valeurs moyennes des poids entre les entrées et les unités de sélection accordée par le module de sélection en fonction du type des entrées (entrées sensorielles, graphe de planification) et de la nature de l'expert (guidage, lieu, Figure 4.6). Elles peuvent indiquer en effet quel type d'entrée va le plus influencer un expert donné.

Par exemple, dans l'expérience de Roberts et Pearce (1999), le blocage des amers distaux par le groupe Trial-Same (qui n'a eu accès aux amers distaux qu'après une exposition avec une balise) n'a pas été aisé à expliquer avec les seuls taux de sélection des experts Planification, ces taux étant similaires à la fin de l'apprentissage dans les groupes Trial-Same et Session-Same (groupe qui avait été exposé à des amers distaux pendant l'entraînement et ne présentait pas de blocage, Figure 8.6 a). En revanche, l'évolution des poids dans le module de sélection se révèle plus informative : le groupe Session-Same montre en effet une montée des poids entre le graphe et l'unité de Planification, alors qu'elle est totalement absente chez le groupe Trial-Same (Figure 8.6 b).

Nombre de paramètres à définir. Nous terminons cette section en notant l'usage pratique qui peut être fait du modèle par le *peu de paramètres* qu'il y a eu à modifier pour reproduire les expériences variées.

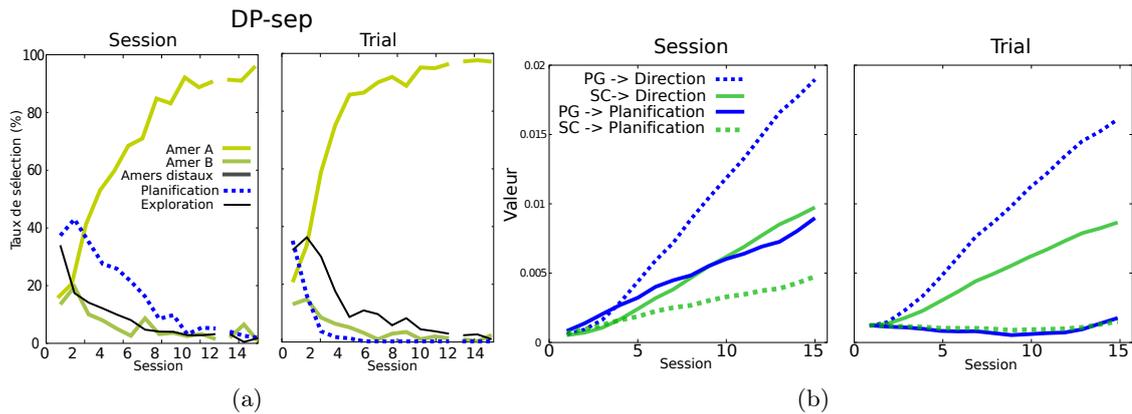


Figure 8.6 – **Expérience de Roberts et Pearce (1999)**. a) Evolution des taux de sélection du groupe DP-sep. b) Evolution des poids synaptiques entre les entrées sensorielles et les unités correspondant aux stratégies Direction et Planification au sein du réseau de sélection. PG->Direction : poids entre le graphe de planification et l'unité de l'expert Direction, SC->Direction : entrées sensorielles vers unité Direction, PG->Planification : Graphe vers unité Planification, SC->Planification : Entrées sensorielles vers unité Direction.

Le nombre de paramètres du modèle est conséquent (11 pour une configuration "normale", i.e., un expert de guidage et un de lieu, cf Tableau B.1), cependant seuls les vitesses d'apprentissage des experts procéduraux et du module de sélection ont du être redéfinis (2 dans cette configuration). Le tableau 8.2 montre à quel point ces modifications ont été peu nombreuses.

Nous avons fait ici l'hypothèse qu'une telle variabilité des paramètres pouvait dépendre de facteurs non pris en compte (motivation, demande de la tâche, etc.). Dans un sens, ce choix est raisonnable, compte tenu des connaissances actuelles et des récents travaux de modélisation qui y ont été consacrés (e.g., Doya, 2002; Luksys et al, 2009), montrant que les changements de taux d'apprentissage d'un modèle (même au cours d'une expérience) pouvaient expliquer les changements de niveaux motivationnels, par exemple de stress pour des souris génétiquement modifiées.

La nécessité de changer ces taux pour certaines simulations pourrait donner lieu également à des investigations neurobiologiques quant aux taux d'ACh dans les différents systèmes neuraux supposés être responsables de ces stratégies (voir Chapitre 2) et confirmer que, par exemple, dans l'expérience de Roberts et Pearce (1999), qui ne nécessite pas de sélections rapides et autorisant ainsi la présence de blocage à l'essai de test, les mesures correspondraient à des taux moins importants d'ACh (paramètre fixé à 0.01) que dans celle de Devan et White (1999) (paramètre fixé à 1), dans laquelle les stratégies de guidage et de lieu doivent être rapidement sélectionnées.

Ce modèle peut donc représenter un outil utile pour l'analyse des comportements de navigation animale. Comme nous l'avons montré dans cette section ainsi que dans les sections précé-

Tableau 8.2 – Paramètres des taux d'apprentissage et de sélection selon les expériences (cf Tableau 8.1 pour les noms et groupes associés)

Expérience	Guidage	Sélection
A-EXP1	0.25	0.25
B-EXP2	0.25	0.25
C-Morris	0.01	0.1
D-Steele	0.25	0.25
E-Rodrigo	0.25	0.25
F-Roberts	0.1	0.1
G-McGregor	0.25	0.25
H-Devan	0.25	1
I-Pearce	0.1	0.25

denes, il offre la possibilité d'anticiper notamment le type de stratégies utilisées, le moment de leur changement ou leurs influences réciproques, anticipations qui pourraient donner lieu à des expérimentations animales les confortant.

Il n'a été cependant été validé que dans un même dispositif, la piscine de Morris, et sur une tâche simple. Même si certains travaux supposent que les interactions entre stratégies ne sont pas semblables dans un autre dispositif (Botreau et Gisquet-Verrier, 2010), il reste encore beaucoup de points à investiguer dans les protocoles que ce labyrinthe permet.

Le dernier Chapitre conclut sur de brèves perspectives concernant à la fois le domaine des neurosciences et de la robotique.

Chapitre 9

Conclusion et Perspectives

Le modèle computationnel bioinspiré de sélection de stratégies de navigation présenté dans ce travail a pour objectif d'apporter une contribution au débat concernant la nature de l'apprentissage spatial chez l'animal, en particulier chez le rat. On rappelle que ce débat concerne la façon dont les amers environnementaux sont pris en compte dans les différents moyens qu'utilisent les animaux pour rejoindre un but, entrant indépendamment en compétition pour contrôler le comportement ou, au contraire, étant intégrés dans une représentation cohérente de l'environnement.

Nous avons constaté l'impossibilité pour les configurations du modèle ne supposant qu'une compétition d'amers (les groupes D-par) ou seulement une intégration (les groupes L-par et P-par) de reproduire correctement les expériences. La grande majorité des expériences a pu être en effet reproduite avec les groupes Contrôle (TL, DL, TP ou DP) avec entrées séparées, dans lesquels différentes stratégies pouvaient être apprises en parallèle sur la base de types d'amers proximaux ou distaux.

Nous en avons tiré des similitudes avec des propositions issues de la théorie associative :

(i) l'hypothèse de combinaison adaptative (Ratcliff et Newcombe, 2008), qui n'exclut pas la prise en compte d'amers indépendants, mais suppose que certains amers ont pu être intégrés par l'animal au cours d'expériences passées et reconnus d'emblée comme étant utiles dans un contexte particulier. Notre modèle inclut cette séparation de manière ad hoc, mais propose une spécialisation de ces types d'amers pour des experts distincts, amers indépendants pour experts de guidage, amers intégrés pour experts de lieu ;

(ii) l'hypothèse du comparateur (Miller et Matzel, 1988), qui propose que certains amers ou groupes d'amers peuvent acquérir une valeur associative plus forte que d'autres sans entraîner un blocage de l'apprentissage de tous les autres amers. Le module de sélection du modèle, qui tient compte à la fois des entrées sensorielles et de l'historique des sélections, prend en compte les différentes stratégies sans en éliminer totalement certaines ;

et des propositions des théories cognitive et modulaire :

(iii) la notion de carte topologique et l'hypothèse d'apprentissages différentiels des amers,

pour que des stratégies soient complémentaires et pertinentes dans des environnements stables ou instables, implantées dans des experts procéduraux et déclaratifs.

Ce modèle peut fournir certaines informations qui sont malaisées, voire impossibles à appréhender dans une expérience biologique, comme le type d'amers sur lesquels se fonde l'orientation courante de l'animal, comment ils sont utilisés pour rejoindre le but, avec quel cadre de référence et quel mode d'apprentissage. Surtout – et cela représente l'originalité de ce modèle – en autorisant la sélection de stratégies aux caractéristiques hétérogènes, il permet de donner un éclairage particulier sur les interactions que peuvent avoir les différents systèmes neuraux sous-tendant chacune de ces stratégies, ainsi que sur le concours de ces interactions dans les effets relatifs à l'un ou l'autre des paradigmes évoqués plus haut.

Nous avons rappelé en Introduction le cadre de l'approche animat dans lequel se place ce travail. Les modèles computationnels issus de cette approche visent, d'une part, à mieux comprendre comment le comportement des animaux participe à leur survie, ou du moins à leur maintenance dans une zone de confort et, d'autre part, à améliorer l'autonomie d'un système artificiel en important les principes modélisés. En l'occurrence, le modèle présenté ici vise à être utile pour l'analyse des expérimentations sur la navigation animale en général – et sur l'apprentissage spatial en particulier – et à être en même temps apte à l'implémentation sur un robot réel, afin de le rendre plus autonome et adaptatif dans un environnement le moins possible préparé par l'Homme.

Quelques travaux déjà engagés et des perspectives vont être avancés dans l'un et l'autre domaine.

9.1 De *Rattus rattus*...

Un certain nombre d'améliorations ont déjà été apportées par rapport aux premières versions du modèle (Dollé et al, 2008, 2010a,b). La prise en compte de stratégies ayant des entrées intégrées ou indépendantes a nécessité leur normalisation, ce qui n'a pas été sans conséquences sur les résultats produits. Par exemple, le graphe de planification a ainsi gagné de l'influence pour le choix des stratégies. Cependant, dans certains cas où les entrées visuelles étaient les plus pertinentes (e.g., dans Pearce et al, 1998), cela a retardé le changement progressif entre stratégies (la mise à jour des poids provenant du graphe de planification étant plus longue) et légèrement affaibli les performances obtenues (sans toutefois invalider les résultats).

De même, l'intégration d'un modèle inspiré par les structures hippocampiques, permettant de remplacer des unités "cellules de lieu" ad hoc par des unités prenant véritablement en compte les amers présents, a élargi considérablement le spectre des protocoles pouvant être reproduits. Cependant nous avons noté que son intégration avait eu des répercussions sur le rôle de l'exploration, beaucoup plus sélectionnée avec des cellules ad hoc. Cela a pu conduire, dans certains cas, à des interprétations divergentes des interactions entre stratégies. Par exemple, dans la reproduction de l'expérience de Devan et White (1999), les jours où le but est caché, les taux de sélection des experts avaient indiqué un rôle majoritaire de l'exploration (Dollé et al, 2010a), alors que dans la reproduction actuelle l'expert de Planification est majoritairement sélectionné, l'exploration étant très minoritaire. Notons cependant que ces derniers résultats sont plus plausibles,

de par le rôle de la stratégie de lieu remarqué par Devan et White (1999).

Dans sa version actuelle, le modèle semble en l'état suffisamment robuste pour avoir pu s'adapter aisément aux diverses conditions environnementales proposées ici, composées d'un nombre varié d'amers proximaux et distaux.

Cependant il n'a pas pris en compte tous les amers que nous avons mentionnés (cf section 2.1), notamment les amers "atypiques", comme la configuration géométrique de l'environnement ou les bords du dispositif. En effet, ils semblent avoir un statut spécial car ils appartiennent au dispositif – donc nettement distinguables des amers –, mais l'animal ne peut pas les contourner, contrairement aux amers proximaux. Or ces amers trouvent un écho dans le débat associatif-cognitif, car ils seraient, selon certains, impénétrables aux effets de compétition associative. L'un de ces échos se retrouve par exemple dans le travail de Doeller et al (2008), distinguant l'apprentissage des "frontières" du dispositif de l'apprentissage des autres amers.

Notre modèle a révélé en effet un traitement spécial à réserver aux bords de la piscine de Morris, notamment dans deux expériences qui lui ont été impossible de reproduire, la première dans laquelle les amers étaient situés sur les bords de la piscine (Chamizo et Rodrigo, 2004), la seconde dans laquelle les animaux avaient une tendance à préférer la localisation absolue de la plateforme, et non la relative, lorsque la piscine était déplacée et ses bords cachés (Hamilton et al, 2008). Un expert prenant en compte les "cellules de bord", récemment découvertes dans l'hippocampe par Solstad et al (2008), pourrait être envisagé. Il serait ainsi classé dans une autre catégorie que les autres, avec des spécificités en terme d'apprentissage et de cadre de référence particulières. Cela ne présume en rien de son impénétrabilité vis-à-vis des autres amers, hypothèse que nous pourrions être en mesure d'investiguer dans la suite de notre travail.

Quant à la gestion de la deuxième catégorie d'amers atypiques, la configuration géométrique de l'environnement, qui est au cœur des dernières avancées de ce débat associatif-cognitif (e.g., Sturz et Bodily, 2010; Ponticorvo et Miglino, 2010), elle reste l'objet de nos réflexions, car les avis sur son existence sont contradictoires. En effet, certains travaux supposent qu'elle est traitée par un module à part entière (e.g., Cheng, 1986; Redhead et Hamilton, 2007), d'autres que ce n'est que l'émergence d'un apprentissage spatial dans un contexte particulier (e.g., Pearce et al, 2006; Sheynikhovich et al, 2009).

Bien d'autres ajouts restent encore à effectuer pour rendre ce modèle plus opérationnel, comme la reconnaissance des amers ou encore l'encodage du but de façon autonome, ce qui nécessite de prendre en compte des états motivationnels, non encore intégrés dans cette architecture. Un modèle computationnel développé dans notre équipe, inspiré des boucles cortico-basales et gérant une sélection de l'action soumise à plusieurs motivations, est en cours d'implémentation (Girard et al, 2005, 2008; Liénard, 2009)¹.

1. Rapport de Jean Liénard téléchargeable sur http://chronos.isir.upmc.fr/~Dolle/lienard_memoire.pdf



Figure 9.1 – Superposition de l’environnement robotique et des préférences de sélection de stratégies apprises par le robot (disques verts et rouge pour, respectivement, les stratégies Taxon et Planification). Le but est indiqué par l’hexagone bleu : le robot a appris à rejoindre le but visible avec Taxon dans son voisinage, et à utiliser la Planification lorsqu’il est hors du champ visuel. Photographie et montage : Ken Caluwaerts.

9.2 ... à Psikharpax

L’implémentation de ce modèle de sélection de stratégies de navigation n’a débuté que très récemment au cours d’un stage dans notre laboratoire (Caluwaerts, 2010)².

La plateforme robotique est le robot-rat Psikharpax, issu du Projet Européen ICEA (Integrating Cognition, Emotion and Autonomy, 2006-2009, e.g., Meyer et al, 2005; NGuyen et al, 2010; Bernard et al, 2010), dont deux prototypes sont véritablement fonctionnels depuis début 2010. Ce robot vise à être contrôlé par des architectures computationnelles fortement inspirées des circuits nerveux du rat.

Le robot sur lequel a été implémenté une première version du modèle ne se sert pour l’instant que de sa vision et de son odométrie, alors qu’il est doté également de systèmes auditif, tactile (ses vibrisses) et vestibulaire. Ses traitements visuels s’effectuent avec une carte de vision bioinspirée (société BVS), avec laquelle il peut détecter des formes, des mouvements et des couleurs.

2. Rapport de Ken Caluwaerts téléchargeable sur http://chronos.isir.upmc.fr/~caluwaerts/thesis_ken_caluwaerts.pdf

Les premières étapes ont révélé la difficulté que le robot a de prendre en compte les amers "naturels", malgré son traitement visuel sophistiqué. L'environnement expérimental est donc pour l'instant largement préparé par l'Homme (exemple dans Figure 9.1), où des amers contrastés ont été placés à divers endroits du dispositif. Notons que ces amers sont plutôt de type distaux, ce qui a été fait à dessein car c'est exclusivement avec eux que le robot s'est forgé une carte de cellules de lieu. Il peut également utiliser une stratégie Taxon, pré-apprise – un module de catégorisation des amers étant en cours de préparation pour permettre son apprentissage en ligne –, dont la performance est volontairement bridée lorsque le robot est éloigné du but.

Les premières expériences ont consisté à tester le module de sélection du modèle en faisant naviguer le robot vers un but de façon à ce que le module choisisse d'abord une stratégie de lieu quand il est loin du but, puis une stratégie Taxon lorsqu'il s'en approche (comme dans, e.g., Gothard et al, 1996; Tamara et al, 2010). Après avoir préalablement construit sa représentation spatiale avec 100 cellules de lieu et appris une stratégie Planification, d'une part (cf. Caluwaerts, 2010), et avoir appris la stratégie Taxon, d'autre part, le robot réussit à sélectionner la stratégie adéquate pendant sa trajectoire, en partant d'endroits différents. La Figure 9.1 illustre le robot s'approchant du but (hexagone bleu au sol) avec une stratégie de guidage, alors qu'il avait adopté auparavant une stratégie de lieu – car il ne pouvait pas voir la cible. Sur la photo, une superposition entre les emplacements du robot et ses préférences pour l'une ou l'autre stratégie montre en effet des disques verts ou rouges de tailles plus ou moins importantes, représentant les emplacements où il a plus ou moins souvent sélectionné une stratégie Taxon (disques verts) ou Planification (disques rouges). Ce changement au cours d'une trajectoire peut ainsi fournir aux robots une autonomie accrue en lui permettant de choisir dès que possible une stratégie moins exigeante en terme de coût de calcul.

Ce début d'implémentation - n'utilisant pas encore la distinction entre amers proximaux et distaux – montre la capacité de ce modèle à fonctionner sans d'importantes modifications dans un support robotique. Les futures expériences et améliorations que l'on fournira à cette architecture de contrôle, en particulier par des interactions étroites avec les collègues neurobiologistes, informaticiens et roboticiens qui, nous l'espérons, prêteront attention à notre travail, permettront à un système artificiel inspiré de mieux naviguer en gérant à bon escient les différents amers à disposition dans l'environnement.

Sixième partie

Annexes

A Table des figures

2.1	Catégorisation des types d'amers et indices environnementaux permettant de localiser un but. Ce but se trouve ici sous la balise.	13
2.2	Référentiels égocentriques et allocentriques	14
2.3	Stratégies de navigation	15
2.4	Stratégie Praxis	16
2.5	Stratégies de navigation	17
2.6	Blocage et occultation	20
2.7	Schéma des structures neurales de l'encéphale du rat impliquées dans la navigation.	25
2.8	Compétition et coopération entre systèmes neuraux	30
3.1	Renforcement dans un espace d'actions continu	42
3.2	Stratégies Locale et Planification	43
3.3	Le modèle de Uchibe et Doya (2005)	46
3.4	Le modèle de Guazzelli et al (1998)	48
3.5	Le modèle de Girard et al (2005)	49
3.6	Le modèle de Daw et al (2005)	50
3.7	Le modèle de Chavarriaga et al (2005).	52
4.1	Modélisation de l'amer et référentiel	59
4.2	Référentiel allocentrique dans une piscine de Morris	60
4.3	Réseau de neurones modélisant la structure des experts de guidage	61
4.4	Modèle d'hippocampe	64
4.5	Structure du graphe de Planification	68
4.6	Réseau de sélection	71
4.7	Vue générale du modèle	72
4.8	Codes couleurs des groupes	74
5.1	L'environnement simulé : une piscine de Morris de 2m de diamètre	79
5.2	EXP1 : L'environnement simulé : piscine de Morris avec but fixe et balise.	80
5.3	EXP1 : apprentissage des groupes et taux de sélection associés	81
5.4	EXP1 : Cartes de planification	82
5.5	EXP2 : L'environnement simulé : piscine de Morris avec but fixe et amer à côté du but.	83
5.6	EXP2 : apprentissage des groupes et taux de sélection associés	84
5.7	McGregor et al (2004a) : Protocole et résultats observés	86

5.8	McGregor et al (2004a), résultats simulés : apprentissage des groupes pendant l'entraînement et performances aux essais test	88
5.9	McGregor et al (2004a), résultats simulés : Taux de sélection	89
5.10	McGregor et al (2004a), résultats simulés : Cartes de navigation des experts Taxon et Direction	90
5.11	Morris et al (1982) : Protocole et résultats observés	94
5.12	Morris et al (1982), résultats simulés : apprentissage des groupes et taux de sélection associés	95
5.13	Influence de la taille de l'amer sur l'apprentissage	96
5.14	Morris et al (1982) : Exemples de trajectoire	97
5.15	Steele et Morris (1999) : résultats observés	99
5.16	Steele et Morris (1999), résultats simulés : apprentissage des groupes	100
5.17	Steele et Morris (1999), résultats simulés : taux de sélection et valeurs de sélection des amers	101
6.1	Rodrigo et al (2006) : protocole et résultats observés	109
6.2	Rodrigo et al (2006), résultats simulés : apprentissage des groupes et taux d'occupation dans les essais tests	111
6.3	Rodrigo et al (2006), résultats simulés : taux de sélection et taux d'occupation dans les tests d'extinction	113
6.4	Rodrigo et al (2006), résultats simulés : taux d'occupation dans les essais tests	114
6.5	Rodrigo et al (2006), résultats simulés : sélections d'experts dans deux trajectoires et cartes de planification associées	116
6.6	Roberts et Pearce (1999) : protocole et résultats observés	119
6.7	Roberts et Pearce (1999), résultats simulés : apprentissage des groupes et temps passé dans le quadrant du but à l'essai de test	122
6.8	Roberts et Pearce (1999), résultats simulés : apprentissage des groupes et temps passé dans le quadrant du but à l'essai de test	123
6.9	Roberts et Pearce (1999), résultats simulés : taux de sélection pendant l'essai de test	124
6.10	Roberts et Pearce (1999), résultats simulés : taux de sélection (DP-sep)	125
6.11	Roberts et Pearce (1999), résultats simulés : comparaison des performances entre groupes Trial et Same	125
6.12	Roberts et Pearce (1999), résultats simulés : taux de sélection (DP-par et D-par)	127
6.13	Roberts et Pearce (1999), résultats simulés : taux de sélection pendant l'essai de test	128
6.14	McGregor et al (2004a) : protocole et résultats observés	132
6.15	McGregor et al (2004a), résultats simulés : rappel des groupes Rotate	133
6.16	McGregor et al (2004a), résultats simulés : apprentissage des groupes et performances dans les essais de test	134
6.17	McGregor et al (2004a), résultats simulés : tau de sélection dans les essais de test (DP-par)	136
6.18	McGregor et al (2004a), résultats simulés : exemple de valeurs de sélection dans une trajectoire	137

6.19	McGregor et al (2004a), résultats simulés : performances dans les essais de test du groupe P-sep	138
6.20	McGregor et al (2004a), résultats simulés : taux de sélection pendant l'entraînement et les essais de test (DP-sep, DL-sep et DL-par)	139
7.1	Devan et White (1999) : protocole et résultats observés	147
7.2	Devan et White (1999), résultats simulés : apprentissage des groupes (TL-sep, TL-par, TP-sep)	148
7.3	Devan et White (1999), résultats simulés : taux de sélection par session, selon le contexte	150
7.4	Devan et White (1999), résultats simulés : Cartes navigationnelles	152
7.5	Devan et White (1999), résultats simulés : taux de sélection par type d'individu ("cue- ou place-responder"	153
7.6	Devan et White (1999), résultats simulés : taux de sélection (TL-par, TP-par)	154
7.7	Devan et White (1999), résultats simulés : taux de sélection selon le protocole	155
7.8	Pearce et al (1998) : protocole et résultats observés	158
7.9	Pearce et al (1998), résultats simulés : apprentissage des groupes (Contrôles vs lésés)	160
7.10	Pearce et al (1998), résultats simulés : apprentissage des groupes (P-sep vs D-sep)	161
7.11	Pearce et al (1998), résultats simulés : apprentissage des groupes (Sessions vs Trial)	162
7.12	Pearce et al (1998), résultats simulés : taux d'occupation et taux de sélection dans les premiers et quatrièmes essais	163
7.13	Pearce et al (1998), résultats simulés : trajectoires typiques et carte de navigation associées	165
7.14	Pearce et al (1998); Devan et White (1999), résultats simulés : Apprentissage des groupes et taux de sélection associés	169
8.1	Influence des apprentissages entre experts : expérience de Devan et White (1999)	178
8.2	Exemple d'une influence des apprentissages entre experts : expérience de Roberts et Pearce (1999)	179
8.3	Expérience de McGregor et al (2004a) prolongée sur 20 sessions	187
8.4	Différentes possibilités d'implémenter une stratégie de guidage avec deux amers	190
8.5	Résultats simulés du modèle dans un labyrinthe en T	194
8.6	Analyse des poids du réseau de sélection	197
9.1	Plateforme robotique Psikharpax	202
C.1	Morris et al (1982) : carte de l'utilisation des amers selon l'emplacement du but	213
D.1	Rodrigo et al (2006) : taux de sélection (TP-sep, TP-par, T-par)	215
D.2	Rodrigo et al (2006) : taux de sélection dans les essais de test	216
D.3	Roberts et Pearce (1999) : temps passé dans le quadrant du but (TP-sep, TP-par, T-par)	217
D.4	Roberts et Pearce (1999) : Taux de sélection des experts par sessions (TP-sep, TP-par, T-par)	218

D.5	Roberts et Pearce (1999) : taux de sélection (DL-sep)	219
D.6	Roberts et Pearce (1999) : temps passé dans le quadrant du but et taux de sélection associés (DL-sep)	219
E.1	Devan et White (1999) : apprentissage des groupes (TP-par, DP-sep, DP-par)	221
E.2	Devan et White (1999) : taux de sélection associés	222
E.3	Pearce et al (1998) : apprentissage des groupes (DP-par)	223
E.4	Pearce et al (1998) : taux de sélection associés (DP-par)	223
F.1	Chamizo et Rodrigo (2004) : protocole et résultats observés	226
F.2	Chamizo et Rodrigo (2004) : cartes de navigation	226
F.3	Représentation des entrées de l'expert "parois"	227
F.4	Hamilton et al (2008) : protocole	228
F.5	Hamilton et al (2008) : protocole et résultats simulés	229

Annexe B : Paramètres du modèle

Tableau B.1 – Paramètres du modèle utilisés pour les expériences. ¹ assigné pour donner une représentation suffisamment détaillée aux experts ; ² adapté de Chavarriaga et al (2005) ; ³ adapté à la main.

Nom	Valeur	Description
Expert Taxon		
N_{SC} ¹	100	Nombre de SC pour chaque amer
N_{AC}^T ²	36	Nombre de cellules d'action
σ^T ²	22.5	Écart-type de profil de généralisation
λ^T ²	0.95	Facteur d'amointrissement de la trace d'éligibilité
γ^T ²	0.8	Facteur d'amointrissement de la récompense future
η^T ³	0.25	Taux d'apprentissage
Expert Direction		
λ^D ²	0.95	Facteur d'amointrissement de la trace d'éligibilité
γ^D ²	0.8	Facteur d'amointrissement de la récompense future
η^D ³	0.25	Taux d'apprentissage
Expert Local		
λ^D ²	0.95	Facteur d'amointrissement de la trace d'éligibilité
γ^D ²	0.8	Facteur d'amointrissement de la récompense future
η^L ³	0.01	Taux d'apprentissage
Expert Planification		
α ³	0.7	Facteur d'amointrissement de la valeur de but
N_{PC} ¹	1000	Nombre de cellules de lieux créées
Réseau de sélection		
γ_G ²	0.8	Facteur d'amointrissement de la récompense future
λ_G ²	0.7	Taux d'amointrissement de la trace d'éligibilité.
ξ ²	0.25	Taux d'apprentissage

Le tableau B.1 résume l'ensemble des paramètres utilisés communément pour les expériences reproduites. Dans la section qui suit, nous indiquons, pour les expériences ayant été simulées avec des paramètres différents, les valeurs que ceux-ci ont alors prises. Nous rappelons également le tableau des paramètres ayant été modifiés (Tableau B.2).

B.1 Paramètres modifiés selon ces expériences

B.1.1 Morris et al (1982)

- η^D , Taux d'apprentissage de l'expert Direction = 0.01
- ξ , Taux d'apprentissage du réseau de sélection = 0.1

B.1.2 Roberts et Pearce (1999)

- η^D , Taux d'apprentissage de l'expert Direction = 0.1
- ξ , Taux d'apprentissage du réseau de sélection = 0.1

B.1.3 Devan et White (1999)

- ξ , Taux d'apprentissage du réseau de sélection = 1

B.1.4 Pearce et al (1998)

- η^D , Taux d'apprentissage de l'expert Direction = 0.1

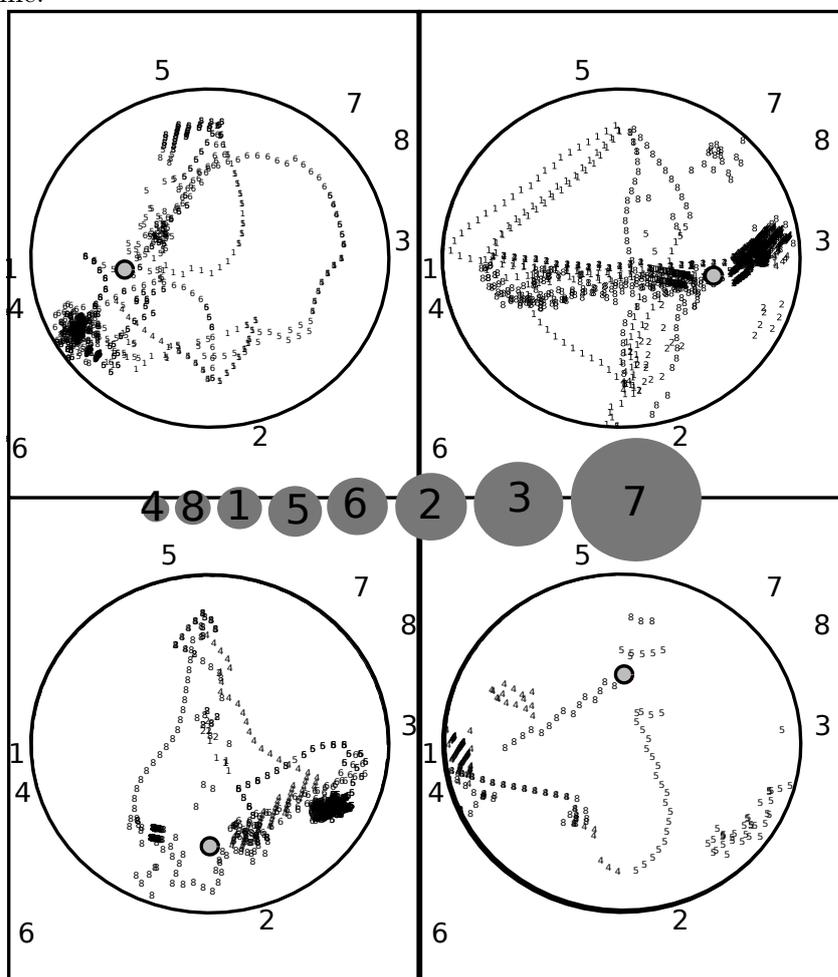
Tableau B.2 – Résumé des paramètres des taux d'apprentissage et de sélection selon les expériences

Expérience	Guidage	Sélection
A-EXP1	0.25	0.25
B-EXP2	0.25	0.25
C-Morris	0.01	0.1
D-Steele	0.25	0.25
E-Rodrigo	0.25	0.25
F-Roberts	0.1	0.1
G-McGregor	0.25	0.25
H-Devan	0.25	1
I-Pearce	0.1	0.25

Annexe C : Expérience de Morris et al (1982)

La Figure C.1 montre le choix des amers distaux par l'expert Direction pour retrouver un but caché. Ce choix n'est pas toujours conditionné par sa position relative au but (e.g., l'utilisation des amers 8 et 4 pour retrouver le but au sud).

Figure C.1 – **Simulation de l'expérience inspirée de celle de Morris et al (1982)**. Groupe D : numéros des amers distaux utilisés par le méta-expert Direction selon la position du but lors d'un dernier essai. De gauche à droite et de haut en bas : emplacements Ouest, Est, Sud, Nord de la plateforme.



Annexe D : Interactions entre amers proximaux ou/et distaux

D.1 Rodrigo et al (2006)

Les taux de sélection des experts des groupes TP-sep, TP-par et T-par (Figure D.1) montrent une sélection majoritaire du Taxon “proximal” dans tous les cas. De même, l’expert Planification n’est utilisé que lorsque le Taxon n’a pas accès à l’amer distal, dans TP-sep. La Figure D.2 permet de mieux évaluer le comportement de l’animat dans les jours de test pour tous les groupes.

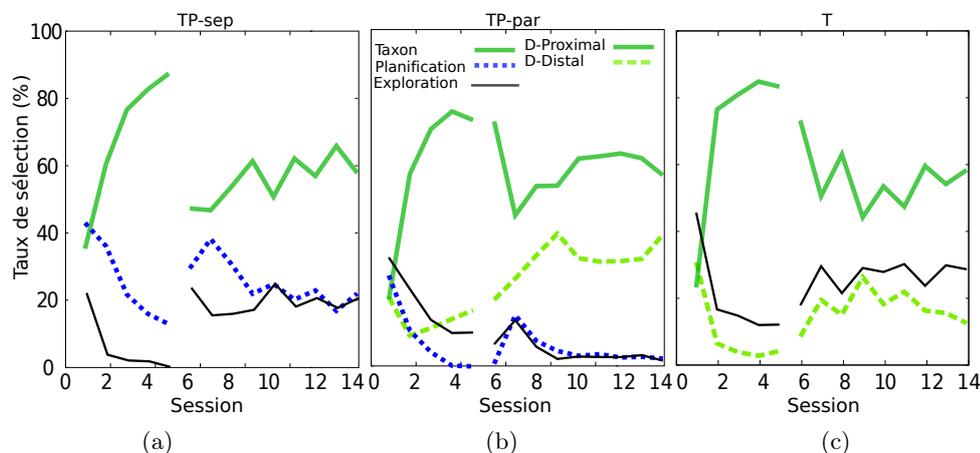


Figure D.1 – Simulation de l’expérience de Rodrigo et al (2006), Sélection des experts dans les groupes TP-sep, TP-par et T-par

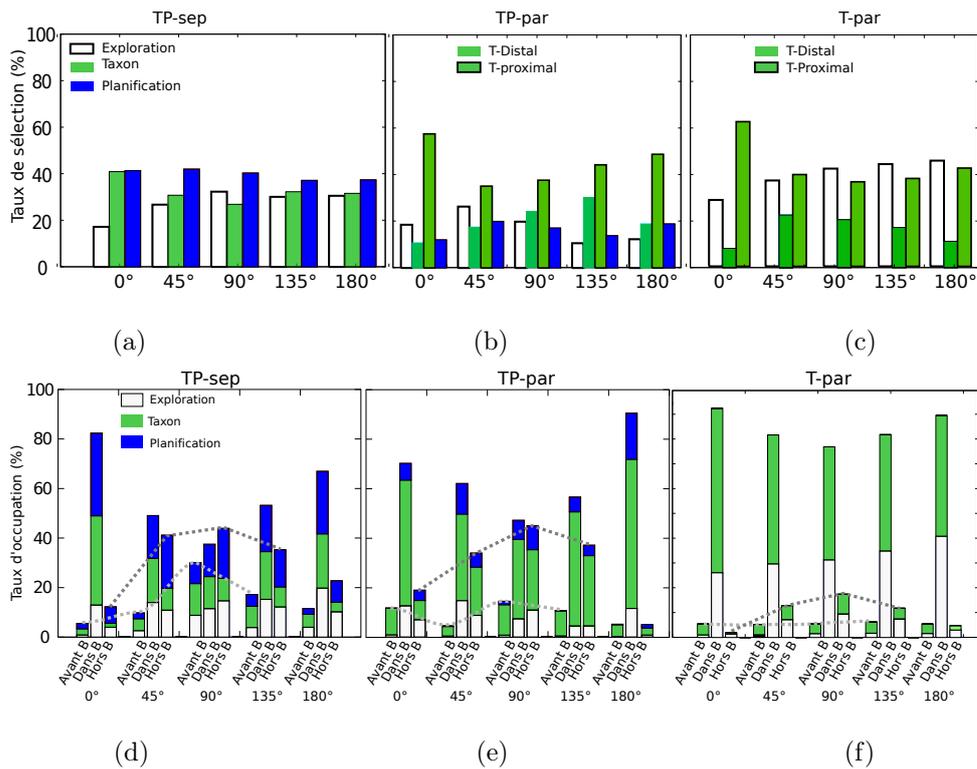


Figure D.2 – **Simulation de l'expérience de Rodrigo et al (2006)**, (a-c.) Taux de sélection des experts pendant les 5 essais test correspondant à chacun des 5 déplacements de l'amer proximal B. (d-f) Taux de sélection des stratégies pondérés par le temps passé pendant les 5 essais test avant d'atteindre l'octant B pour la première fois et après, dans cet octant, et hors de cet octant.

D.2 Expérience de Roberts et Pearce (1999)

D.2.1 Groupes avec Taxon

Les groupes avec Taxon (TP-sep, TP-par, T-par) ne reproduisent pas les résultats attendus : le groupe TP-sep ne montre pas de blocage (Figure D.3 a) tandis que les deux autres groupes montrent du blocage dans tous les cas (Figure D.3 b et c). Les sélections du groupe TP-sep à l'essai de test (Figure D.3 d) montrent une plus forte sélection de l'expert Planification par rapport au groupe TP-par (Figures D.3 e).

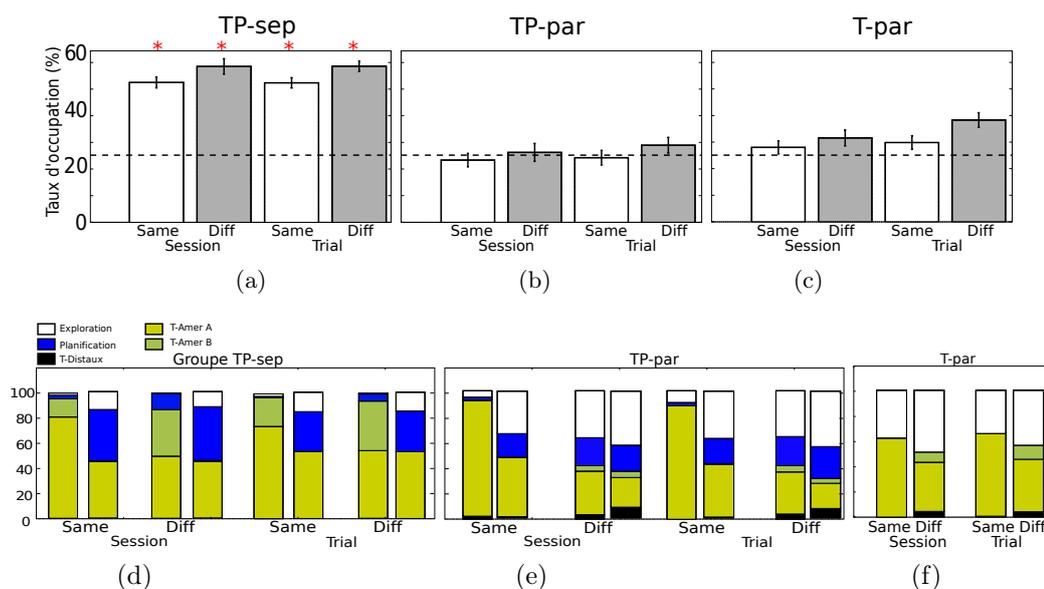


Figure D.3 – **Simulation de l'expérience de Roberts et Pearce (1999)**. (a-c) Temps passé par les groupes TP-sep, TP-par et T-par dans le quadrant du but pendant l'essai de test. Les astérisques noires indiquent les différences significatives entre groupes ; les astérisques rouges indiquent que le résultat du groupe est significativement différent du hasard. (d-f) Sélection des différents experts *juste avant* l'essai de test (première barre de chaque groupe) et *pendant* l'essai de test (deuxième barre).

Les sélections de ces groupes au cours de l'entraînement montrent en effet que le groupe TP-sep recrute les experts Taxon A et B (Figure D.4 a), ce qui permet à l'expert Planification d'être plus facilement sélectionné à l'essai test. Les deux autres groupes montrent une sélection quasi-exclusive de l'expert Taxon A (Figure D.4 b et c), ce qui explique le blocage de ces groupes pendant l'essai de test.

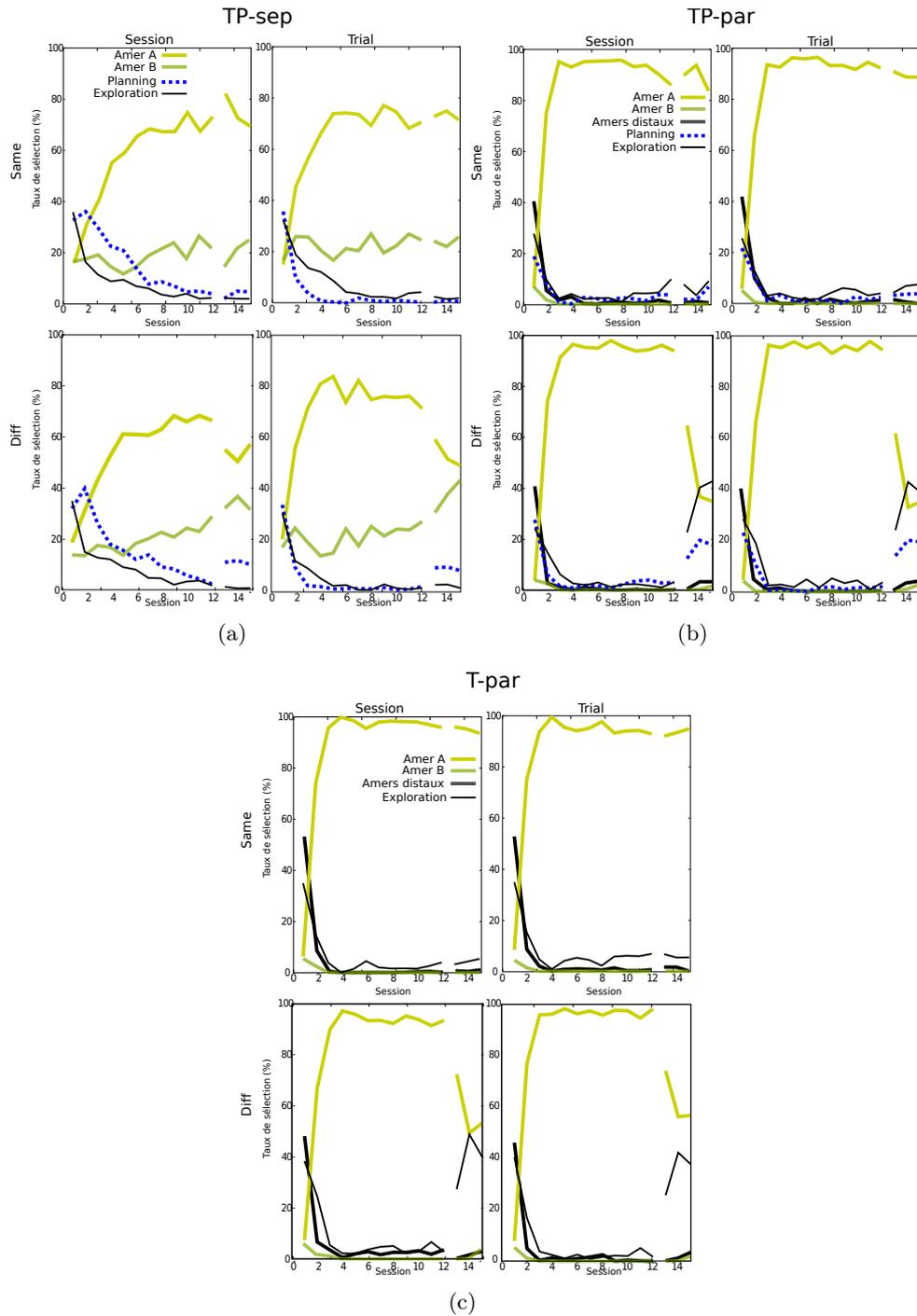


Figure D.4 – Simulation de l'expérience de Roberts et Pearce (1999). Taux de sélection par sessions des groupes (a) TP-sep, (b) TP-par, (c) T-par.

D.2.2 Groupe DL-sep

Le groupe DL-sep montre dans tous les groupes une sélection de l'expert Local légèrement supérieure par rapport aux autres (Figure D.5). L'absence de blocage dans tous les groupes (Figure D.6 a) est confirmée par la forte sélection de l'expert Local, qui a appris les 8 emplacements possibles de la plateforme, le jour de test, dans tous les groupes (Figure D.6 b).

On indique en discussion que l'apprentissage réciproque entre les experts Local et Taxon est responsable de cette absence de blocage.

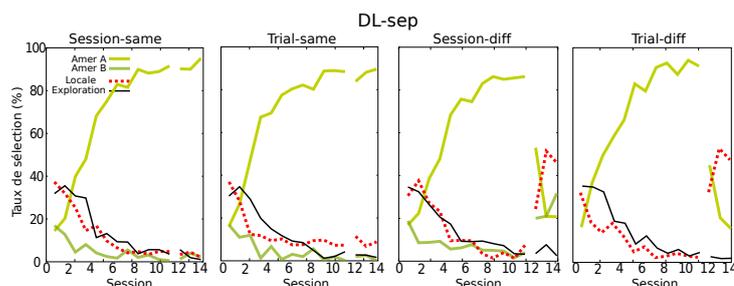


Figure D.5 – **Simulation de l'expérience de Roberts et Pearce (1999)**. Sélection des experts dans les groupes Session-Same, Trial-Same, Session-diff et Trial-diff.

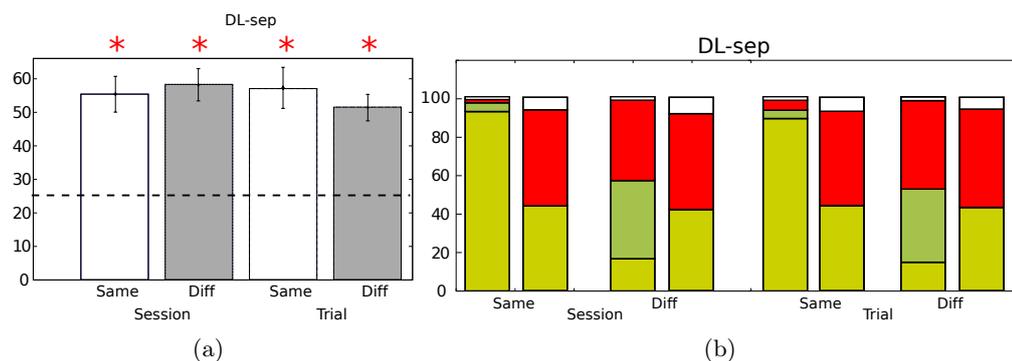


Figure D.6 – **Simulation de l'expérience de Roberts et Pearce (1999)**. (a) Temps passé dans le quadrant du but pendant l'essai de test. Les astérisques noirs sont les différences significatives entre groupes ; les astérisques rouges indiquent que le résultat du groupe est significativement différent du hasard. (b) Sélection des différents experts juste avant l'essai de test (première barre de chaque groupe) et pendant l'essai de test (deuxième barre).

Annexe E : Interactions entre systèmes de guidage et de lieu en situation de conflit

E.1 Expérience de Devan et White (1999)

Lors du test, les groupes T-par, D-sep et D-par (Figure E.1 a, b et c) montrent des taux d'échappement plus haut qu'ils ne devraient. Pour le groupe T-par, cela est dû à l'influence des amers proximaux (comme le montrent les taux de sélection du groupe TP-par, Figure E.2 b, on remarque le même effet pour le groupe TL-par, Figure E.2 a). Pour les groupes DP-sep et D-sep, la nature allocentrique de l'expert Direction ne lui permet pas de trouver directement le but visible comme le pourrait un expert Taxon, d'où la forte sélection d'exploration (Figures E.2 c, e et f). Enfin, le groupe DP-par montre la plus forte influence des amers distaux (Figure E.2 d) qui prennent de plus en plus d'importance au cours de l'entraînement.

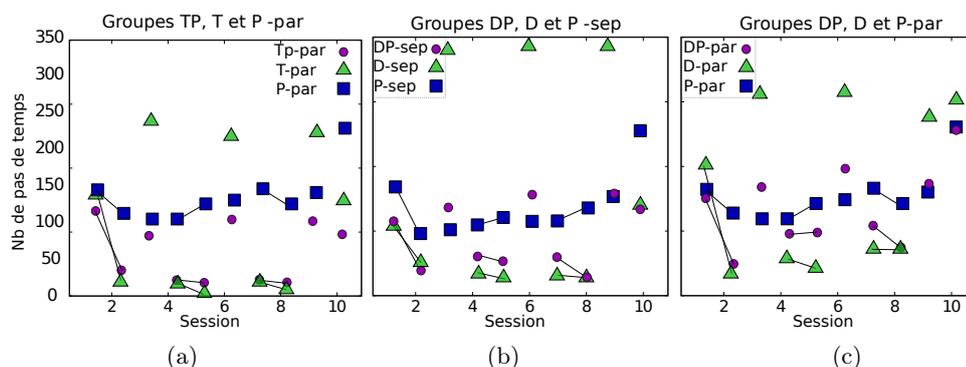


Figure E.1 – **Simulation de l'expérience de Devan et White (1999)**. (a-f) Résultats obtenus en simulation avec (a) avec les groupes TP,T,P-par ; (b) avec les groupes DP,D,P-sep ; (c) avec les groupes DP,D,P-par.

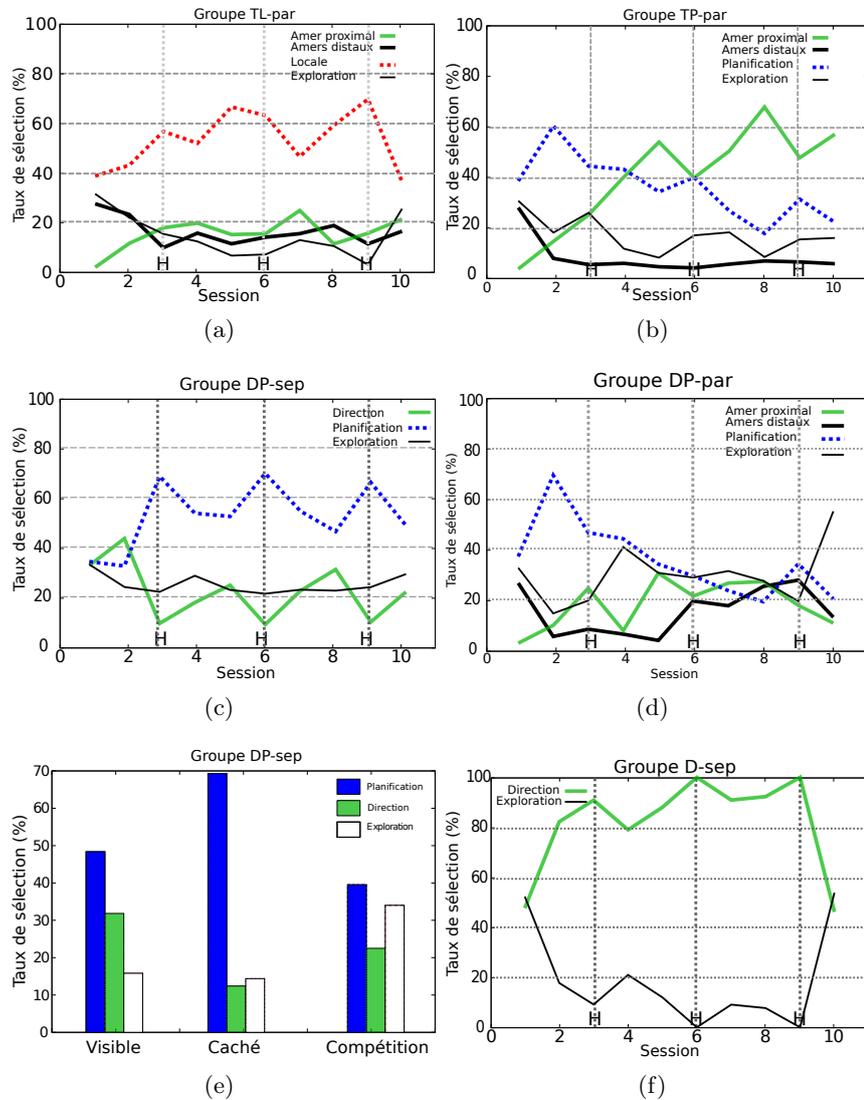


Figure E.2 – **Simulation de l'expérience de Devan et White (1999)** Taux de sélection des experts par session dans les groupes (a) TL-par, (b) TP-par, (c) DP-sep, (d) DP-par, (f) D-sep. (e) Taux de sélection cumulés des experts dans le groupe DP-sep, selon la configuration de l'environnement (but visible, caché et test de compétition).

E.2 Expérience de Pearce et al (1998)

L'influence des amers distaux dans le groupe DP-par est telle qu'il n'y a pas de différence entre les apprentissages inter-sessions de ce groupe et le groupe associé P-par (Figures E.3 a et b). Les sélections aux 4^{èmes} essais dans les groupes DP-par et D-par (Figures E.4 a-d) montrent que ces amers distaux sont plus sélectionnés que l'amer proximal dans le groupe DP-par, tandis que leur sélection n'augmente que légèrement dans le groupe D-par.

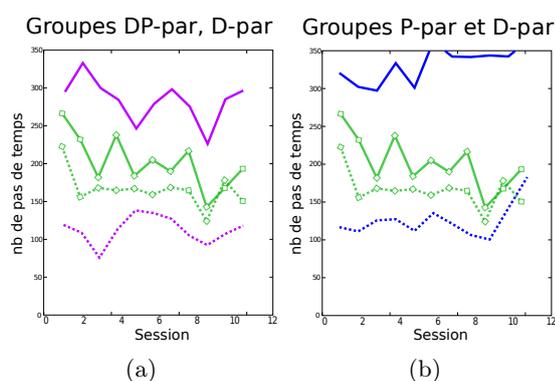


Figure E.3 – **Expérience de Pearce et al (1998)** (a-b) Performances des groupes DP-par/D-par et D-par/P-par.

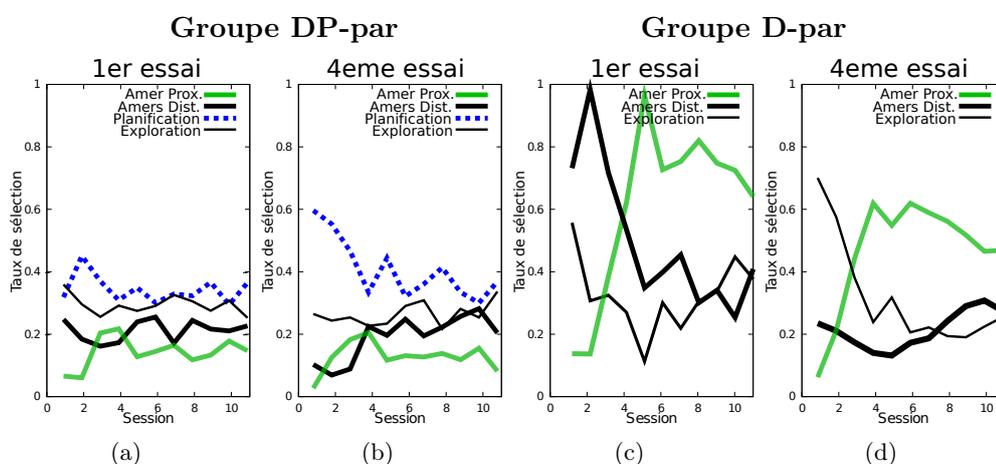


Figure E.4 – **Simulation de l'expérience de Pearce et al (1998)** (a-b) Groupe Contrôle DP-par; (c-d) Groupe D-par : taux de sélection par session des experts dans les premiers et quatrièmes essais.

Annexe F : Expériences supplémentaires

Nous présentons ici les résultats de simulations n'ayant pu reproduire ceux observés.

F.1 Expérience de Chamizo et Rodrigo (2004) : limitation des stratégies de guidage

La vaine tentative de reproduire l'expérience de Chamizo et Rodrigo (2004), dans laquelle les rat apprennent à rejoindre une plateforme en fonction d'amers situés sur les murs de la piscine, l'un proche et l'autre plus éloigné du but (Figure F.1 a), nous amène aussi à questionner la modélisation de nos stratégies de guidage (Taxon et Direction).

Les animaux apprennent plus facilement à retrouver le but s'il est proche d'un amer (30 cm) qu'éloigné (60 cm, Figure F.1 b). Cependant nous obtenons des performances inversées et surtout globalement médiocres (Figure F.1 c). La stratégie Direction, mais aussi Taxon, montrent donc des limites quant à sa possibilité d'inférer la position du but par rapport à sa proximité à un amer. Les cartes de navigation dans les deux cas (Figure F.2) montrent en effet que l'expert a appris des stratégies de contournement de l'amer, passant majoritairement vers le but, mais limitées par la taille de l'environnement (i.e., certaines actions proposées proposent d'aller dans le mur). Il est surtout impossible de déterminer de ces cartes une stratégie mieux ou moins bien apprise que l'autre.

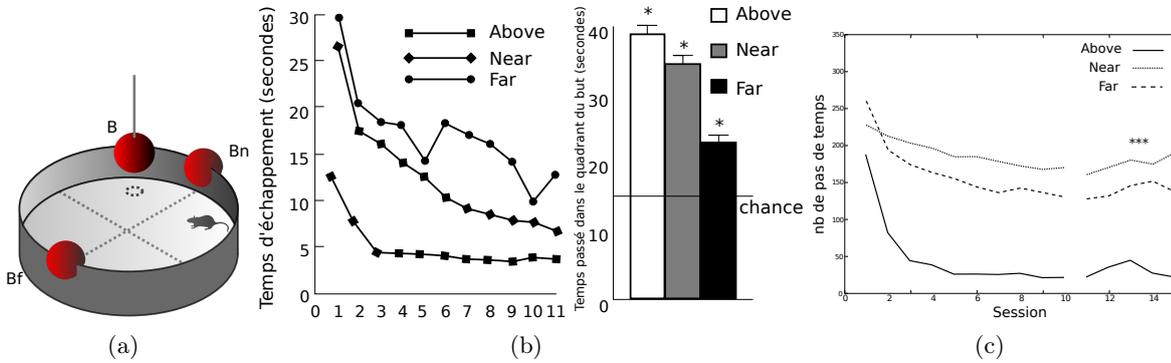


Figure F.1 – a) Protocole de l’expérience de Chamizo et Rodrigo (2004) : trois groupes apprennent à rejoindre une plateforme selon sa position par rapport à un amer, qui peut -être au-dessus de la plateforme (B, groupe “Above”), à côté (Bn, groupe “Near”) ou éloigné (Bf, groupe “Far”) b) Résultats originaux de l’expérience de Chamizo et Rodrigo (2004), c) Résultats simulés avec le groupe DP-sep.

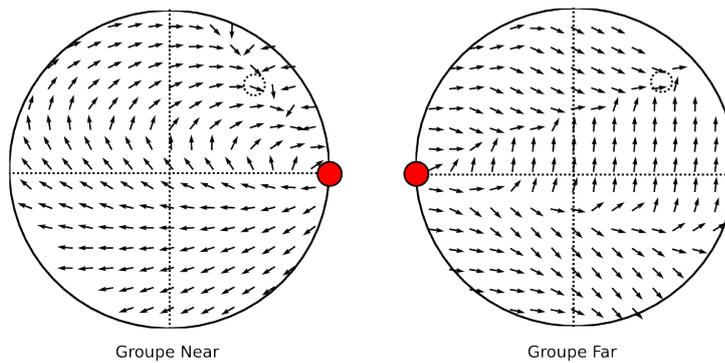


Figure F.2 – Cartes de navigation de l’expert Direction dans les conditions “Near” et “Far”.

F.2 Expert Parois

Inspirés par l'article de Doeller et al (2008) suggérant la nature déclarative de l'apprentissage de tels amers, nous avons modélisé une première version de cet expert "parois" comme un expert de lieu, se forgeant une carte constituée de cellules qui s'activent d'autant plus fortement que l'animal est proche du mur (Figure F.3). Ces cellules pouvaient représenter les "cellules de bord", récemment découvertes dans l'hippocampe par Solstad et al (2008). Sans obtenir toutefois des résultats complètement probants, l'ajout d'un tel expert au modèle a suffisamment amélioré le comportement de l'animat dans l'expérience d'Hamilton et al. pour laisser penser que, modifié, il pourrait être nécessaire, non seulement pour reproduire ces expériences, mais pour reconsidérer les simulations effectuées auparavant.

F.3 Expérience de Hamilton et al (2008)

Nous avons évoqué précédemment que les raisons pour lesquelles les experts de lieu et de guidage semblaient parfois limités pouvaient venir de l'absence de certaines stratégies non implémentées. Nous prenons l'exemple d'un expert parois, dont la conception est inspirée par le rôle que peuvent avoir les limites de l'environnement, comme l'a noté Hamilton et al (2008), dont le protocole est montré sur la figure F.4.

Pendant l'entraînement (Figure F.4 a), l'animal reçoit deux types d'information pour localiser le but : la direction générale indiquée par les amers distaux et l'emplacement absolu selon sa carte cognitive. Lors des essais de test, pour différencier ces deux types d'informations, les auteurs déplacent alors la piscine selon une translation (Figure F.4 b), afin que l'emplacement du but selon la direction générale (relatif) et l'emplacement absolu selon la carte cognitive soient discriminables. Le temps passé dans le voisinage de chacun des emplacements est alors enregistré. Les résultats montrent une préférence significative des animaux pour l'emplacement relatif, suggérant l'utilisation par les animaux d'une stratégie visuelle de type Taxon plutôt que la carte cognitive. Toutefois, en réitérant l'expérience avec la piscine remplie à ras bord (Figure F.4 bas, annulant ainsi la vision de la piscine comme indice potentiel), les auteurs trouvent une

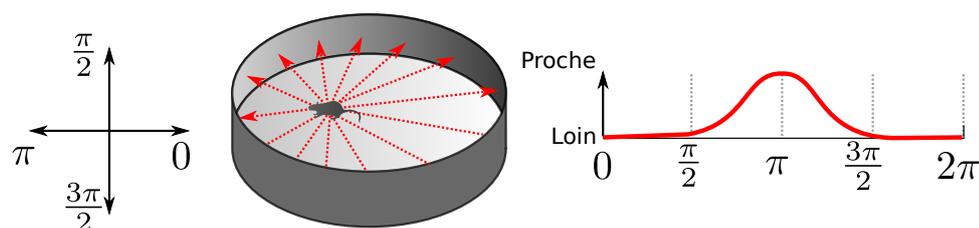


Figure F.3 – Représentation des entrées de l'expert "parois" utilisé dans l'expérience de Hamilton et al (2008) : Les distances de l'animat au mur sont calculées tous les $\pi/50$ (100 données pour 2π , milieu), et la représentation finale (droite) consiste en un vecteur de 100 neurones dont l'activation est d'autant plus forte que le mur est proche de l'animat dans la direction donnée.

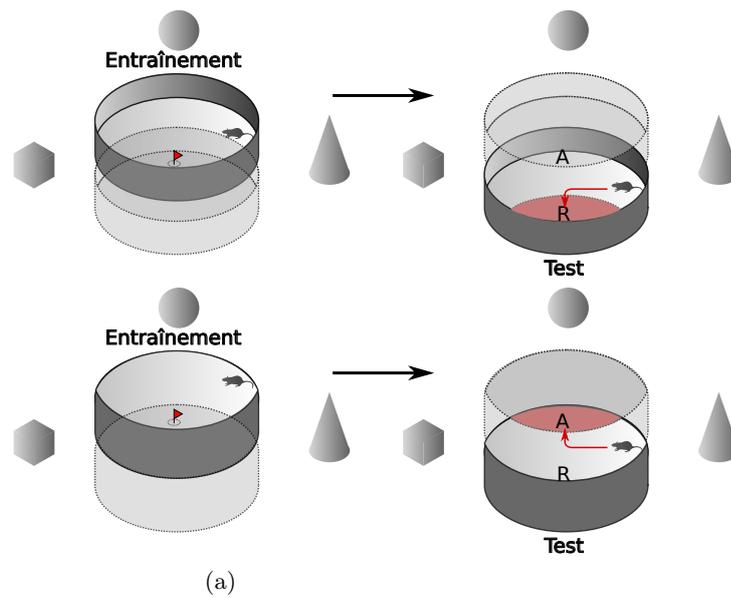


Figure F.4 – Protocole de l’expérience de Hamilton et al (2008). a) Entraînement : la piscine est situé au Nord de l’environnement. b) Test : la piscine est décalée au Sud, recoupant la moitié de la surface auparavant utilisée (zone A)

préférence, plus modeste que la première, pour l’emplacement absolu.

Cet expert parois n’a pas été utilisé pour les expériences présentées ici, du fait de son échec à reproduire les résultats de l’expérience de Hamilton et al (2008), mais l’effet qu’il a eu sur le comportement de l’animat dans cette simulation laissent penser qu’un tel expert pourrait être amélioré pour finalement obtenir le résultat voulu (Figure F.5 a et b).

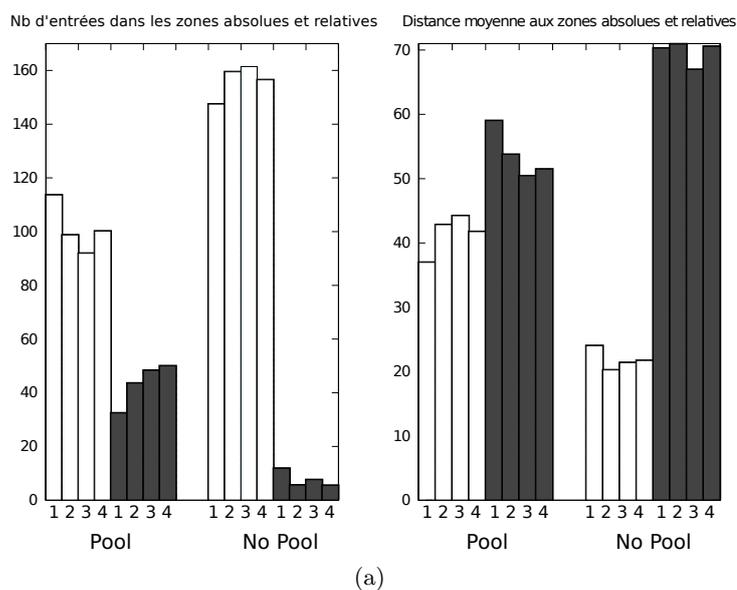


Figure F.5 – Résultats de la simulation : bien que l'animat préfère largement la position absolue, a) il visite plus souvent l'emplacement relatif en présence de l'expert Limite (condition "Pool") qu'en son absence (condition "No pool"), b) sa distance moyenne par rapport à la zone relative est plus faible en condition "Pool" qu'en condition "No Pool".

Références

- Arbib M, Lieblein I (1977) Motivational learning of spatial behaviors. Dans : Metzler J (ed) *Systems Neuroscience*, Academic Press, pp 221–239
- Arkin R (1998) *Behavior-based robotics*. The MIT Press
- Arleo A, Gerstner W (2000) Spatial cognition and neuro-mimetic navigation : a model of hippocampal place cell activity. *Biol Cybern* 83(3) :287–299
- Arleo A, Rondi-Reig L (2007) Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms. *J Integr Neurosci* 6(3) :327–366
- Arleo A, Smeraldi F, Hug S, Gerstner W (2001) Place cells and spatial navigation based on vision, path integration, and reinforcement learning. Dans : *Advances in Neural Information Processing Systems* 13, vol 13, pp 89–95
- Ashby W (1952) *Design for a brain*. Chapman & Hall 100 :49–53
- Atallah H, Lopez-Paniagua D, Rudy J, O'Reilly R (2006) Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nat Neurosci* 10 :126–131
- Avni R, Zadicario P, Eilam D (2006) Exploration in a dark open field : a shift from directional to positional progression and a proposed model of acquiring spatial information. *Behav Brain Res* 171(2) :313–323
- Barlow J (1964) Inertial navigation as a basis for animal navigation. *J Theor Biol* 6(1) :76–117
- Barrera A, Weitzenfeld A (2008) Biologically-inspired robot spatial cognition based on rat neurophysiological studies. *Auton Robots* 25 :147–169
- Batty E, Hoban L, Spetch M, Dickson C (2009) Rats' use of geometric, featural and orientation cues to locate a hidden goal. *Behav process* 82(3) :327–334
- Bennett A (1996) Do animals have cognitive maps? *J Exp Biol* 199(1) :219–224
- Bernard M, N'Guyen S, Pirim P, Gas B, Meyer JA (2010) Phonotaxis behavior in the artificial rat *psikharpax*. Dans : *International Symposium on Robotics and Intelligent Sensors, IRIS2010*, Nagoya, Japon

- Berthoz A (1991) Reference frames for the perception and control of movement. Dans : Paillard J (ed) *Brain and Space*, Oxford University Press, USA, pp 81–111
- Biegler R, Morris R (1993) Landmark stability is a prerequisite for spatial but not discrimination learning. *Nature* 361(6413) :631–633
- Biegler R, Morris R (1999) Blocking in the spatial domain with arrays of discrete landmarks. *J Exp Psychol : Anim Behav Process* 25(3) :334–351
- Birrell JM, Brown VJ (2000) Medial frontal cortex mediates perceptual attentional set shifting in the rat. *J Neurosci* 20(11) :4320–4324
- Blaisdell A (2009) The role of associative processes in spatial, temporal, and causal cognition. Dans : Watanabe SB, Blaisdell AP, Huber L, Young A (eds) *Rational animals, irrational humans*, Tokyo : Keio University Press, pp 153–172
- Blodgett H (1929) The effect of the introduction of reward upon the maze performance of rats. University of California publications in psychology
- Botreau F, Gisquet-Verrier P (2010) Re-thinking the role of the dorsal striatum in egocentric/response strategy. *Front Behav Neurosci* 4 :1–12
- Boulougouris V, Dalley J, Robbins T (2007) Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behav Brain Res* 179(2) :219–228
- Brasted P, Humby T, Dunnett S, Robbins T (1997) Unilateral lesions of the dorsal striatum in rats disrupt responding in egocentric space. *J Neurosci* 17(22) :8919–8926
- Broadbent N, Squire L, Clark R (2007) Rats depend on habit memory for discrimination learning and retention. *Learn Mem* 14(3) :145–151
- Brooks R (1990) Elephants don't play chess. *Rob Auton Syst* 6(1-2) :3–15
- Brown J, Bullock D, Grossberg S (1999) How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J Neurosci* 19(23) :10,502–10,511
- Brown VJ, Bowman EM (2002) Rodent models of prefrontal cortical function. *Trends Neurosci* 25(7) :340–343
- Burguière E (2006) Rôle du cervelet dans la navigation : étude du mécanisme cellulaire de dépression synaptique à long terme des fibres parallèles. Rapport de Thèse, Université Pierre et Marie Curie Paris VI
- Burnod Y (1991) Organizational levels of the cerebral cortex : an integrated model. *Acta Biotheor* 39(3-4) :351–361
- Busch M, Skubic M, Keller J, Stone K (2007) A robot in a water maze : learning a spatial memory task. Dans : 2007 IEEE International Conference on Robotics and Automation, pp 1727–1732

- Caluwaerts K (2010) Design of a biologically inspired navigation system for the psikharpax rodent robot. Mémoire de Master, Université Pierre et Marie Curie – Universiteit Gant
- Carillo-Mora P, Magda G, Abel S (2009) Spatial memory : Theoretical basis and comparative review on experimental methods in rodents. *Behav Brain Res* 203(2) :151–164
- Cartwright B, Collett T (1983) Landmark learning in bees. *J Comp Physiol Sensory Neural Behav Physiol* 151(4) :521–543
- Chakravarthy V, Joseph D, Bapi R (In Press) What do the basal ganglia do? A modeling perspective. *Biol Cybern* pp 1–17
- Chamizo V (2003) Acquisition of knowledge about spatial location : Assessing the generality of the mechanism of learning. *Q J Exp Psychol* 56(1) :102–113
- Chamizo V, Rodrigo T (2004) Effect of absolute spatial proximity between a landmark and a goal. *Learn Motiv* 35(2) :102–114
- Chamizo V, Sterio D, Mackintosh N (1983) Blocking and overshadowing between intra-maze and extra-maze cues : A test of the independence of locale and guidance learning. *Q J Exp Psychol B* 37(3) :235–253
- Chang Q, Gold PE (2003) Switching memory systems during learning : changes in patterns of brain acetylcholine release in the hippocampus and striatum in rats. *J Neurosci* 23(7) :3001–3005
- Chapuis N (1987) Detour and shortcut abilities in several species of mammals. Dans : NATO Advanced Study Institutes series D, Behavioural and social sciences, Kluwer, 36, pp 97–106
- Chavarriaga R, Strössl T, Sheynikhovich D, Gerstner W (2005) A computational model of parallel navigation systems in rodents. *Neuroinformatics* 3(3) :223–242
- Cheng K (1986) A purely geometric module in the rat's spatial representation. *Cognition* 23(2) :149–178
- Cheng K, Shettleworth S, Huttenlocher J, Rieser J (2007) Bayesian integration of spatial information. *Psychol Bull* 133(4) :625–638
- Clark R, Broadbent N, Squire L (2007) The hippocampus and spatial memory : Findings with a novel modification of the water maze. *J Neurosci* 27(25) :6647–6654
- Collett T (1992) Landmark learning and guidance in insects. *Phil Trans Biol Sci* 337(1281) :295–303
- Collett T, Cartwright B, Smith B (1986) Landmark learning and visuo-spatial memories in gerbils. *J Comp Physiol Sensory Neural Behav Physiol* 158(6) :835–851
- Conde F, Maire-Lepoivre E, Audinat E, Crepel F (1995) Afferent connections of the medial frontal cortex of the rat. II. Cortical and subcortical afferents. *J Comp Neurol* 352(4) :567–593

- Cuperlier N, Quoy M, Gaussier P (2007) Neurobiologically inspired mobile robot navigation and planning. *Front in Neurorobot* 1 :1–15
- Damper R, French R, Scutt T (2000) ARBIB : An autonomous robot based on inspirations from biology. *Rob Auton Syst* 31(4) :247–274
- Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorso-lateral striatal systems for behavioral control. *Nat Neurosci* 8(12) :1704–1711
- De Bruin J, Swinkels W, De Brabander J (1997) Response learning of rats in a morris water maze : involvement of the medial prefrontal cortex. *Behav Brain Res* 85(1) :47–55
- De Leonibus E, Oliverio A, Mele A (2005) A study on the role of the dorsal striatum and the nucleus accumbens in allocentric and egocentric spatial memory consolidation. *Learn Mem* 12(5) :491–503
- DeCoteau W, Kesner R (2000) A Double Dissociation Between the Rat Hippocampus and Medial Caudoputamen in Processing Two Forms of Knowledge. *Behav Neurosci* 114(6) :1096–1108
- Denève S, Latham P, Pouget A (1999) Reading population codes : a neural implementation of ideal observers. *Nat Neurosci* 2(8) :740–745
- Denniston J, Savastano H, Miller R (2001) The extended comparator hypothesis : Learning by contiguity, responding by relative strength. Dans : Mowrer RR, Klein S (eds) *Handbook of contemporary learning theories*, vol 3, Lawrence Erlbaum Associates, pp 65–117
- Devan B, White N (1999) Parallel information processing in the dorsal striatum : relation to hippocampal function. *J Neurosci* 19(7) :2789–2798
- D’Hooge R, De Deyn P (2001) Applications of the morris water maze in the study of learning and memory. *Brain Res Rev* 36(1) :60–90
- Diez-Chamizo V, Sterio D, Mackintosh N (1985) Blocking and overshadowing between intra-maze and extra-maze cues : A test of the independence of locale and guidance learning. *Q J Exp Psychol B* 37(3) :235–253
- Dijkstra E (1959) A note on two problems in connection with graphs. *Numer Math* 1(269–270) :269–271
- Doeller CF, Burgess N (2008) Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proc Natl Acad Sci U S A* 105(15) :5909–5914
- Doeller CF, King JA, Burgess N (2008) Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc Natl Acad Sci U S A* 105(15) :5915–5920
- Dollé L, Khamassi M, Girard B, Guillot A, Chavarriaga R (2008) Analyzing interactions between navigation strategies using a computational model of action selection. Dans : *Spatial Cognition VI*, Springer-Verlag, LNAI 5248, vol 5248, pp 71–86

- Dollé L, Sheynikhovich D, Girard B, Chavarriaga R, Guillot A (2010a) Path planning versus cue responding : a bioinspired model of switching between navigation strategies. *Biol Cybern* 103(4) :299–317
- Dollé L, Sheynikhovich D, Girard B, Ujfallussy B, Chavarriaga R, Guillot A (2010b) Analyzing interactions between cue-guided and place-based navigation with a computational model of action selection : Influence of sensory cues and training. Dans : Doncieux S, Girard B, Guillot A, Hallam J, Meyer J, Mouret J (eds) *Proceedings of The Eleventh International Conference on Simulation of Adaptive Behavior*, vol 6226, pp 335–346
- Doya K (2002) Metalearning and neuromodulation. *Neural Net* 15(4-6) :495–506
- Doya K, Samejima K, Katagiri, K KM (2002) Multiple model-based reinforcement learning. *Neural Comput* 14(6) :1347–1369
- Drai D, Kafkafi N, Benjamini Y, Elmer G, Golani I (2001) Rats and mice share common ethologically relevant parameters of exploratory behavior. *Behav Brain Res* 125(1-2) :133–140
- Eichenbaum H, Yonelinas A, Ranganath C (2007) The medial temporal lobe and recognition memory. *Annual Review of Neuroscience* 30 :123–152
- Eilam D, Golani I (1989) Home base behavior of rats (*Rattus norvegicus*) exploring a novel environment. *Behav Brain Res* 34(3) :199–211
- Etienne A, Jeffery K (2004) Path integration in mammals. *Hippocampus* 14(2) :180–192
- Filliat D, Meyer JA (2002) Global localization and topological map-learning for robot navigation. Dans : *From Animals to Animats 7 Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, pp 131–140
- Filliat D, Meyer JA (2003) Map-based navigation in mobile robots - I. A review of localisation strategies. *J Cogn Syst Res* 4(4) :243–282
- Foster DJ, Morris RG, Dayan P (2000) A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* 10(1) :1–16
- Franz MO, Mallot HA (2000) Biomimetic robot navigation. *Rob Auton Syst* 30(1) :133–153
- Gallistel C (1990) *The organization of learning*. MIT press Cambridge, MA
- Gaussier P, Revel A, Banquet J, Babeau V (2002) From view cells and place cells to cognitive map learning : processing stages of the hippocampal system. *Biol Cybern* 86(1) :15–28
- Gaussier P, Banquet J, Sargolini F, Giovannangeli C, Save E, Poucet B (2007) A model of grid cells involving extra hippocampal path integration, and the hippocampal loop. *J Integr Neurosci* 6(3) :447–476
- Gibson B, Shettleworth S (2003) Competition among spatial cues in a naturalistic food-carrying task. *Learn Behav* 31(2) :143–159

- Gibson B, Shettleworth S (2005) Place versus response learning revisited : tests of blocking on the radial maze. *Behav Neurosci* 119(2) :567–586
- Gibson J (1966) *The senses considered as perceptual systems*. Houghton Mifflin Boston
- Girard B, Filliat D, Meyer JA, Berthoz A, Guillot A (2005) Integration of navigation and action selection functionalities in a computational model of cortico-basal-thalamo-cortical loops. *Adapt Behav* 13(2) :115–130
- Girard B, Tabareau N, Pham Q, Berthoz A, Slotine J (2008) Where neuroscience and dynamic system theory meet autonomous robotics : a contracting basal ganglia model for action selection. *Neural Net* 21(4) :628–641
- Gold P (2004) Coordination of multiple memory systems. *Neurobiol Learn Mem* 82(3) :230–242
- Goodale M, Foreman N, Milner A (1978) Visual orientation in the rat : A dissociation of deficits following cortical and collicular lesions. *Exp Brain Res* 31(3) :445–457
- Gothard K, Skaggs W, Moore K, McNaughton B (1996) Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *J Neurosci* 16(2) :823
- Gould K, Kelly D, Kamil A (2010) What scatter-hoarding animals have taught us about small-scale navigation. *Phil Trans Biol Sci* 365(1542) :901
- Gouteux S, Thinus-Blanc C, Vauclair J (2001) Rhesus monkeys use geometric and nongeometric information during a reorientation task. *J Exp Psychol Gen* 130(3) :505–519
- Grahn J, Parkinson J, Owen A (2008) The cognitive functions of the caudate nucleus. *Prog Neurobiol* 86(3) :141–155
- Granon S, Poucet B (1995) Medial prefrontal lesions in the rat and spatial navigation : evidence for impaired planning. *Behav Neurosci* 109(3) :474–484
- Groenewegen H, Berendse H, Meredith G, Haber S, Voorn P, Wolters J, Lohman A (1991) The mesolimbic dopamine system : From motivation to action. Dans : Willner P, Scheel-Kriiger J (eds) *Functional anatomy of the ventral, limbic system-innervated striatum*, John Wiley, Chichester., pp 19–59
- Guazzelli A, Corbacho F, Bota M, Arbib M (1998) Affordances, motivation, and the world graph theory. *Adapt Behav* 6(3) :435–471
- Guillot A, Meyer JA (2000) From SAB94 to SAB2000 : What's New, Animat ? Dans : *From animals to animats 6 : proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, The MIT Press, p 3
- Gurney K, Prescott T, Redgrave P (2001a) A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol Cybern* 84(6) :401–410
- Gurney K, Prescott T, Redgrave P (2001b) A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biol Cybern* 84(6) :411–423

- Hafting T, Fyhn M, Molden S, Moser MB, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436(7052) :801–806
- Hamilton D, Rosenfelt C, Whishaw I (2004) Sequential control of navigation by locale and taxon cues in the morris water task. *Behav Brain Res* 154(2) :385–397
- Hamilton D, Akers K, Weisend M, Sutherland R (2007) How do room and apparatus cues control navigation in the morris water task? evidence for distinct contributions to a movement vector. *J Exp Psychol : Anim Behav Process* 33(2) :100–114
- Hamilton D, Akers K, Johnson T, Rice J, Candelaria F, Sutherland R, Weisend M, Redhead E (2008) The relative influence of place and direction in the morris water task. *J Exp Psychol : Anim Behav Process* 34(1) :31–53
- Hamilton D, Akers K, Johnson T, Rice J, Candelaria F, Redhead E (2009) Evidence for a shift from place navigation to directional responding in one variant of the Morris water task. *J Exp Psychol Anim Behav Process* 35(2) :271–278
- Hartley T, Burgess N (2005) Complementary memory systems : competition, cooperation and compensation. *Trends Neurosci* 28(4) :169–170
- Haruno M, Kawato M (2006) Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops : fmri examination in stimulus-action-reward association learning. *Neural Net* 19 :1242–1254
- Harvey D, Brant L, Commins S (2009) Differences in cue-dependent spatial navigation may be revealed by in-depth swimming analysis. *Behav Processes* 82(2) :190–197
- Hasselmo ME (2005) A model of prefrontal cortical mechanisms for goal-directed behavior. *J Cogn Neurosci* 17(7) :1115–1129
- Hayward A, McGregor A, Good M, Pearce J (2003) Absence of overshadowing and blocking between landmarks and the geometric cues provided by the shape of a test arena. *Q J Exp Psychol* 56(1) :114–126
- Horne M, Pearce J (2009) A landmark blocks searching for a hidden platform in an environment with a distinctive shape after extended pretraining. *Learn Behav* 37(2) :167–178
- Houk J, Adams J, Barto A (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. Dans : Davis J (ed) *Models of information processing in the basal ganglia*, The MIT Press, pp 249–270
- Hull C (1943) *Principles of behavior : An introduction to behavior theory*. Appleton-Century-Crofts New York
- Jankowski J, Scheef L, Hüppe C, Boecker H (2009) Distinct striatal regions for planning and executing novel and automated movement sequences. *Neuroimage* 44(4) :1369–1379

- Kamin L (1969) Predictability, surprise, attention, and conditioning. Dans : Campbell B, Church R (eds) *Punishment and Aversive Behavior*, New York : Appleton-Century-Crofts, pp 279–296
- Kaufman M, Bolles R (1981) A nonassociative aspect of overshadowing. *Bull Psychonomic Soc* 18(6) :318–320
- Kealy J, Diviney M, Kehoe E, McGonagle V, O’Shea A, Harvey D, Commins S (2008) The effects of overtraining in the Morris water maze on allocentric and egocentric learning strategies in rats. *Behav Brain Res* 192(2) :259–263
- Kelly D, Gibson B (2007) Spatial navigation : spatial learning in real and virtual environments. *Comput Cogn Behav Rev* 2 :111–124
- Kesner RP, Farnsworth G, DiMattia BV (1989) Double dissociation of egocentric and allocentric space following medial prefrontal and parietal cortex lesions in the rat. *Behav Neurosci* 103(5) :956–961
- Khamassi M (2007) Complementary roles of the rat prefrontal cortex and striatum in reward-based learning and shifting navigation strategies. Rapport de Thèse, Université Paris 6
- Khamassi M, Lachèze L, Girard B, Berthoz A, Guillot A (2005) Actor-critic models of reinforcement learning in the basal ganglia : from natural to artificial rats. special issue towards artificial rodents. *Adapt Behav, Special Issue Towards Artificial Rodents* 13(2) :131–148
- Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex* 13(4) :400–408
- Kim J, Baxter M (2001) Multiple brain-memory systems : the whole does not equal the sum of its parts. *Trends Neurosci* 24(6) :324–30
- Kim J, Ragozzino M (2005) The involvement of the orbitofrontal cortex in learning under changing task contingencies. *Neurobiol Learn Mem* 83(2) :125–133
- Klatzky R (1998) Allocentric and egocentric spatial representations : definitions, distinctions, and interconnections. Dans : Freksa C, Habel C, Wender K (eds) *Spatial cognition. An interdisciplinary approach to representing and processing spatial knowledge*, Springer, pp 107–127
- Kolb B, Walkey J (1987) Behavioural and anatomical studies of the posterior parietal cortex in the rat. *Behav Brain Res* 23(2) :127–145
- Kubie J, Fenton A (2009) Heading-vector navigation based on head-direction cells and path integration. *Hippocampus* 19(5)
- Lagarde M (2010) Apprentissage de nouveaux comportements : vers le développement épigénétique d’un robot autonome. Rapport de Thèse, Université de Cergy-Pontoise
- Lambrinos D, Kobayashi H, Pfeifer R, Maris M, Labhart T, Wehner R (1997) An autonomous agent navigating with a polarized light compass. *Adapt Behav* 6(1) :131–161

- Lavenex P, Schenk F (1998) Olfactory traces and spatial learning in rats. *Anim Behav* 56(5) :1129–1136
- Lee A, Duman R, Pittenger C (2008) A double dissociation revealing bidirectional competition between striatum and hippocampus during learning. *Proc Natl Acad Sci U S A* 105(44) :17,163
- Lehmann H, Clark BJ, Whishaw IQ (2007) Similar development of cued and learned home bases in control and hippocampal-damaged rats in an open field exploratory task. *Hippocampus* 17(5) :370–380
- Leising K, Blaisdell A (2009) Associative basis of landmark learning and integration in vertebrates. *Comp Cogn Behav Rev* 4 :80–102
- Liénard J (2009) Mise au point d'un modele neuromimetique de selection de strategie de navigation en environnement simule. Mémoire de Master, Université Pierre et Marie Curie, Paris 6, France.
- Luksys G, Gerstner W, Sandi C (2009) Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning. *Nat Neurosci* 12(9) :1180–1186
- Maaswinkel H, Whishaw I (1999) Homing with locale, taxon, and dead reckoning strategies by foraging rats : sensory hierarchy in spatial navigation. *Behav Brain Res* 99(2) :143–52
- Mackintosh N (1975) A theory of attention : Variations in the associability of stimuli with reinforcement. *Psychol Rev* 82(4) :276–298
- Martel G, Blanchard J, Mons N, Gastambide F, Micheau J, Guillou J (2007) Dynamic interplays between memory systems depend on practice : The hippocampus is not always the first to provide solution. *Neuroscience* 150(4) :743–753
- Martinet LE, Passot JB, Fouque B, Meyer JA, Arleo A (2008) Map-based spatial navigation : A cortical column model for action planning. *LNAI* 5248 :39–55
- McDonald R, White N (1993) A triple dissociation of memory systems : hippocampus, amygdala, and dorsal striatum. *Behav Neurosci* 107(1) :3–22
- McDonald R, White N (1994) Parallel information processing in the water maze : evidence for independent memory systems involving dorsal striatum and hippocampus. *Behav Neural Biol* 61(3) :260–70
- McDonald R, Devan B, Hong N (2004a) Multiple memory systems : The power of interactions. *Neurobiol Learn Mem* 82(3) :333–346
- McDonald RJ, Hong NS, Devan BD (2004b) The challenges of understanding mammalian cognition and memory-based behaviours : an interactive learning and memory systems approach. *Neurosci Biobehav Rev* 28(7) :719–745
- McFarland D, Sibly R (1975) The behavioural final common path. *Phil Trans Roy Soc Lond B Biol Sci* 270(907) :265–293

- McGregor A, Good M, Pearce J (2004a) Absence of an Interaction Between Navigational Strategies Based on Local and Distal Landmarks. *J Exp Psychol : Anim Behav Process* 30(1) :34–44
- McGregor A, Hayward AJ, Pearce JM, Good MA (2004b) Hippocampal lesions disrupt navigation based on the shape of the environment. *Behav Neurosci* 118(5) :1011–1021
- McIntyre C, Marriott L, Gold P (2003) Patterns of brain acetylcholine release predict individual differences in preferred learning strategies in rats. *Neurobiol Learn Mem* 79(2) :177–183
- Meyer JA, Filliat D (2003) Map-based navigation in mobile robots - II. a review of map-learning and path-planning strategies. *J Cogn Syst Res* 4(4) :283–317
- Meyer JA, Guillot A (1991) Simulation of adaptive behavior in animats : Review and prospect. Dans : *Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats*, MIT Press, pp 2–14
- Meyer JA, Guillot A, Girard B, Khamassi M, Pirim P, Berthoz A (2005) The Psikharpax project : towards building an artificial rat. *Rob Auton Syst* 50(4) :211–223
- Micheau J, Riedel G, Roloff E, Inglis J, Morris R (2004) Reversible Hippocampal Inactivation Partially Dissociates How and Where to Search in the Water Maze. *Behav Neurosci* 118(5) :1022–1032
- Miller JD E K and Cohen (2001) An integrative theory of prefrontal function. *Annu Rev Neurosci* 24(1) :167–202
- Miller N, Shettleworth S (2007) Learning about environmental geometry : An associative model. *J Exp Psychol Anim Behav Process* 33(3) :191–212
- Miller R, Matzel L (1988) The comparator hypothesis : A response rule for the expression of associations. *Psychol Learn Motiv* 22 :51–92
- Miller R, Barnet R, Grahame N (1995) Assessment of the Rescorla-Wagner model. *Psychol Bull* 117(3) :363–386
- Mishkin M, Malamut B, Bachevalier J (1984) Memories and habits : Two neural systems. Dans : Lynch G, McGaugh J, Weinberger M (eds) *Neurobiology of Learning and Memory*, New York : Guilford, pp 65–77
- Mittelstaedt H (2000) Triple-loop model of path control by head direction and place cells. *Biol Cybern* 83(3) :261–270
- Mittelstaedt H, Mittelstaedt M, Papi F, Wallraff H (1982) *Avian navigation*. Springer Berlin
- Mizumori S (2008) *Hippocampal Place Fields*. Oxford University Press, USA
- Montague P, Dayan P, Person C, Sejnowski T (1995) Bee foraging in uncertain environments using predictive hebbian learning. *Nature* 377(6551) :725–728

- Montague P, Hyman S, Cohen J (2004) Computational roles for dopamine in behavioural control. *Nature* 431 :760–767
- Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J Neurosci* 16(5) :1936–1947
- Morris R (1981) Spatial localisation does not require the presence of local cues. *Learn Motiv* 12 :239–260
- Morris R, Garrud J P and Rawlins, O’Keefe J (1982) Place navigation impaired in rats with hippocampal lesions. *Nature* 297(5868) :681–683
- Mountcastle VB (1957) Modality and topographic properties of single neurons of cat’s somatic sensory cortex. *J Neurophysiol* 20(4) :408–434
- Mountcastle VB (1997) The columnar organization of the neocortex. *Brain* 120 (Pt 4) :701–722
- Newcombe N, Huttenlocher J (2006) Development of spatial cognition. Dans : Damon W, Lerner R, Kuhn D, Siegler R (eds) *Handbook of child psychology*, Wiley, pp 734–776
- Newcombe N, Ratliff K (2007) Explaining the development of spatial reorientation : Modularity-plus-language versus the emergence of adaptive combination. Dans : Plumert J, Spencer J (eds) *The emerging spatial mind*, Oxford University Press, pp 53–76
- NGuyen S, Pirim P, Meyer J, Girard B (2010) An integrated neuromimetic model of the saccadic eye movement for the psikharpax robots. Dans : Doncieux S, Girard B, Guilot A, Hallam J, Meyer J, Mouret J (eds) *Proceedings of The Eleventh International Conference on Simulation of Adaptive Behavior*, vol 6226, pp 114–125
- O’Keefe J, Nadel L (1978) *The hippocampus as a cognitive map*. Oxford University Press.
- Oliveira MGM, Bueno OFA, Pomarico AC, Gugliano EB (1997) Strategies used by hippocampal- and caudate-putamen-lesioned rats in a learning task. *Neurobiol Learn Mem* 68(1) :32–41
- Oswald CJ, Good M (2000) The effects of combined lesions of the subicular complex and the entorhinal cortex on two forms of spatial navigation in the water maze. *Behav Neurosci* 114(1) :211–217
- Packard M, McGaugh J (1992) Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks : further evidence for multiple memory systems. *Behav Neurosci* 106(3) :439–446
- Packard M, McGaugh J (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol Learn Mem* 65(1) :65–72
- Pavlov I (1927) *Conditioned reflexes*. London : Oxford University Press
- Pearce J (2009) The 36th Sir Frederick Bartlett Lecture : An associative analysis of spatial learning. *Q J Exp Psychol* 62(1) :1665–1684

- Pearce J, Bouton M (2001) Theories of associative learning in animals. *Ann Rev Psychol* 52(1) :111–139
- Pearce J, Roberts A, Good M (1998) Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. *Nature* 396(6706) :75–77
- Pearce J, Ward-Robinson J, Good M, Fussell C, Aydin A (2001) Influence of a beacon on spatial learning based on the shape of the test environment. *J Exp Psychol Anim Behav Process* 27(4) :329–44
- Pearce J, Graham M, Good M, Jones P, McGregor A (2006) Potentiation, overshadowing, and blocking of spatial learning based on the shape of the environment. *J Exp Psychol* 32(3) :201–214
- Poldrack R, Packard M (2003) Competition among multiple memory systems : converging evidence from animal and human brain studies. *Neuropsychol* 41(3) :245–51
- Ponticorvo M, Miglino O (2010) Encoding geometric and non-geometric information : a study with evolved agents. *Anim Cogn* 13(1) :157–174
- Poucet B (1993) Spatial cognitive maps in animals : New hypotheses on their structure and neural mechanisms. *Psychol Rev* 100(2) :163–182
- Poucet B, Save E (2009) Spatial cognitive maps. Dans : Squire L (ed) *Encyclopedia of Neuroscience*, Elsevier, pp 181–186
- Pych J, Chang Q, Colon-Rivera C, Haag R, Gold P (2005) Acetylcholine release in the hippocampus and striatum during place and response training. *Learn Mem* 12(6) :564–572
- Quoy M, Laroque P, Gaussier P (2002) Learning and motivational couplings promote smarter behaviors of an animat in an unknown world. *Rob Auton Syst* 38(3-4) :149–156
- Ragozzino M, Choi D (2004) Dynamic Changes in Acetylcholine Output in the Medial Striatum During Place Reversal Learning. *Learn Mem* 11(1) :70–77
- Ragozzino M, Detrick S, Kesner R (1999) Involvement of the prelimbic-infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. *J Neurosci* 19(11) :4585–4594
- Ranck JB (1984) Head direction cells in the deep cell layer of dorsal presubiculum in freely moving rats. *Soc Neurosci Abstr* 10(176.12)
- Ratliff K, Newcombe N (2008) Reorienting When Cues Conflict. *Psychol Sci* 19(12) :1301–1307
- Redgrave P, Prescott T, Gurney K (1999) The basal ganglia : a vertebrate solution to the selection problem? *Neuroscience* 89 :1009–1024
- Redhead E, Hamilton D (2007) Interaction between locale and taxon strategies in human spatial learning. *Learn Motiv* 38(3) :262–283

- Redhead E, Roberts A, Good M, Pearce J (1997) Interaction between piloting and beacon homing by rats in a swimming pool. *J Exp Psychol : Anim Behav Process* 23(3) :340–350
- Renaudineau S, Poucet B, Save E (2007) Flexible use of proximal objects and distal cues by hippocampal place cells. *Hippocampus* 17(5) :381–395
- Rescorla R, Wagner A (1972) A theory of pavlovian conditioning : The effectiveness of reinforcement and non-reinforcement. Dans : Black A, Prokasy W (eds) *Classical conditioning II : Current research and theory*, New York : Appleton-Century-Crofts, pp 64–69
- Rich E, Shapiro M (2007) Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *J Neurosci* 27(17) :4747–4757
- Rich E, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. *J Neurosci* 29(22) :7208–7219
- Roberts A, Pearce J (1998) Control of spatial behavior by an unstable landmark. *J Exp Psychol* 24(2) :172–184
- Roberts A, Pearce J (1999) Blocking in the Morris swimming pool. *J Exp Psychol : Anim Behav Proc* 25(2) :225–235
- Rodrigo T, Chamizo V, McLaren I, Mackintosh N (1997) Blocking in the spatial domain. *J Exp Psychol Anim Behav Process* 23 :110–118
- Rodrigo T, Sansa J, Baradad P, Chamizo V (2006) Generalization gradients in a navigation task with rats. *Learn Motiv* 37(3) :247–268
- Rondi-Reig L, Petit G, Tobin C, Tonegawa S, Mariani J, Berthoz A (2006) Impaired sequential egocentric and allocentric memories in forebrain-specific-NMDA receptor knock-out mice during a new task dissociating strategies of navigation. *J Neurosci* 26(15) :4071–4081
- Saksida L, Raymond S, Touretzky D (1997) Shaping robot behavior using principles from instrumental conditioning. *Rob Auton Syst* 22(3-4) :231–249
- Samsonovich A, McNaughton B (1997) Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* 17(15) :5900–5920
- Samu D, Erős P, Ujfalussy B, Kiss T (2009) Robust path integration in the entorhinal grid cell system with hippocampal feed-back. *Biol Cybern* 101(1) :19–34
- Save B E and Poucet (2000) Hippocampal-parietal cortical interactions in spatial cognition. *Hippocampus* 10(4) :491–499
- Save E, Moghaddam M (1996) Effects of lesions of the associative parietal cortex on the acquisition and use of spatial memory in egocentric and allocentric navigation tasks in the rat. *Behav Neurosci* 110(1) :74–85

- Save E, Hok V, Renaudineau S, Parron C, Poucet B (2008) Cue and Goal Encoding in Rodents : A Source of Inspiration for Robotics ? Dans : Jefferies M, Yeap K (eds) *Robotics and cognitive approaches to spatial mapping*, Springer, pp 163–180
- Schultz W (1998) Predictive reward signal of dopamine neurons. *J Neurophysiol* 80(1) :1–27
- Schultz W (2002) Getting formal with dopamine and reward. *Neuron* 36(2) :241–263
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(14) :1593–1599
- Shapiro M, Tanila H, Eichenbaum H (1997) Cues that hippocampal place cells encode : dynamic and hierarchical representation of local and distal stimuli. *Hippocampus* 7(6) :624–642
- Sharp P, Blair H, Etkin D, Tzanetos D (1995) Influences of vestibular and visual motion information on the spatial firing patterns of hippocampal place cells. *J Neurosci* 15(1) :173–189
- Sheynikhovich D, Chavarriaga R, Strösslin T, Arleo A, Gerstner W (2009) Is there a geometric module for spatial orientation ? Insights from a rodent navigation model. *Psychol Rev* 116(3) :540–566
- Sheynikhovich D, Dollé L, Chavarriaga R, Arleo A (2010) Minimal model of strategy switching in the plus-maze navigation tasks. Dans : Doncieux S, Girard B, Guilot A, Hallam J, Meyer J, Mouret J (eds) *Proceedings of The Eleventh International Conference on Simulation of Adaptive Behavior*, vol 6226, pp 390–401
- Skinner B (1938) *The behavior of organisms*. New York : Appleton-Century-Crofts
- Skinner DM, Etchegary CM, Ekert-Maret EC, Baker CJ, Harley CW, Evans JH, Martin GM (2003) An analysis of response, direction, and place learning in an open field and t mazes. *J Exp Psychol Anim Behav Process* 29 :3–13
- Solstad T, Boccara C, Kropff E, Moser M, Moser E (2008) Representation of geometric borders in the entorhinal cortex. *Science* 322(5909) :1865
- Squire L, Zola S (1996) Structure and function of declarative and nondeclarative memory systems. *Proc Natl Acad Sci* 93(24) :13,515–13,522
- Squire L, Stark C, Clark R (2004) The medial temporal lobe. *Annu Rev Neurosci* 27 :279–306
- Sridharan D, Prashanth P, Chakravarthy V (2006) The role of the basal ganglia in exploration in a neural model based on reinforcement learning. *Int J Neural Syst* 16(2) :111–124
- Stahlman W, Blaisdell A (2009) Blocking of spatial control by landmarks in rats. *Behav Proc* 81(1) :114–118
- Stalnaker T, Franz T, Singh T, Schoenbaum G (2007) Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron* 54(1) :51–58

- Steele R, Morris R (1999) Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP 5. *Hippocampus* 9(2) :118–136
- Stout S, Miller R (2007) Sometimes-competing retrieval (SOCR) : A formalization of the comparator hypothesis. *Psychol Rev* 114(3) :759–782
- Strösslin T, Sheynikhovich D, Chavarriaga R, Gerstner W (2005) Robust self-localisation and navigation based on hippocampal place cells. *Neural Net* 18(9) :1125–1140
- Stuchlik J A and Bures (2002) Relative contribution of allothetic and idiothetic navigation to place avoidance on stable and rotating arenas in darkness. *Behav Brain Res* 128(2) :179–188
- Sturz B, Bodily K (2010) Encoding of variability of landmark-based spatial information. *Psychol Res* 74 :1–8
- Sutherland RJ, Hamilton DA (2004) Rodent spatial navigation : at the crossroads of cognition and movement. *Neurosci Biobehav Rev* 28(7) :687–697
- Sutton R (1988) Learning to predict by the methods of temporal differences. *Mach Learn* 3(1) :9–44
- Sutton R (1991) Planning by incremental dynamic programming. Dans : *Proceedings of the Eighth International Workshop on Machine Learning*, Citeseer, pp 353–357
- Sutton R, Barto A (1998) *Reinforcement learning : an introduction*. The MIT Press, Bradford Book
- Tamara C, Leffel J, Timberlake W (2010) Egocentric and allocentric search : effects of platform distance and environmental cues. *Anim Cogn* 13(3) :565–581
- Taube JS, Muller RU, Ranck Jr JB (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J Neurosci* 10(2) :420–435
- Thinus-Blanc C (1996) *Animal spatial cognition : behavioural and brain approach*. World Scientific Pub Co Inc
- Thorndike E (1911) *Animal intelligence : experimental studies*. The Macmillan company
- Timberlake W, Sinning SA, Leffel JK (2007) Beacon training in a water maze can facilitate and compete with subsequent room cue learning in rats. *J Exp Psychol : Anim Behav Process* 33(3) :225–243
- Tolman EC (1948) Cognitive maps in rats and men. *Psychol Rev* 55(4) :189–208
- Tolman EC, Honzik CH (1930) Insight in rats. *University of California Publication in Psychology* 4(14) :215–232
- Trullier O, Wiener SI, Berthoz A, Meyer JA (1997) Biologically-based artificial navigation systems : review and prospects. *Prog Neurobiol* 83(3) :271–285

- Uchibe E, Doya K (2005) Reinforcement learning with multiple heterogeneous modules : A framework for developmental robot learning. Dans : The 4th International Conference on Development and Learning, IEEE Computer Society Press, pp 87–92
- Ujfalussy B, Eros P, Somogyvari Z, Kiss T (2008) Episodes in space : A modelling study of hippocampal place representation. LNAI 5040 :123–136
- Verschure P (1998) Distributed adaptive control : explorations in robotics and the biology of learning. *Informatik/Informatique* 1 :25–29
- Waller D, Lippa Y (2007) Landmarks as beacons and associative cues : Their role in route learning. *Mem Cognit* 35(5) :910–924
- Watson J (1907) Kinaesthetic and organic sensations : Their role in the reactions of the white rat to the maze. *Psychological Review Monograph Supplement* 8(33) :1–100
- Watson J (1948) *Psychology as the Behaviorist views it (1913)*. Readings in the history of psychology New York : Appleton-Century-Crofts
- Webb B (2001) Can robots make good models of biological behaviour? *Behav Brain Sci* 24(6) :1033–1050
- Webb B (2009) Animals versus animats : Or why not model the real iguana? *Adapt behav* 17(4) :269–286
- Whishaw I (1991) Latent learning in a swimming pool place task by rats : Evidence for the use of associative and not cognitive mapping processes. *Q J Exp Psychol B* 43(1) :83–103
- Whishaw IQ (1998) Spatial mapping takes time. *Hippocampus* 8(2) :122–130
- White N (2004) The role of stimulus ambiguity and movement in spatial navigation : a multiple memory systems analysis of location discrimination. *Neurobiol Learn Mem* 82 :216–229
- White N (2005) How independent are parallel memory systems? a theoretical comment on gibson and shettleworth (2005). *Behav Neurosci* 119(4) :1158–64
- White N (2009) Some highlights of research on the effects of caudate nucleus lesions over the past 200 years. *Behav Brain Res* 199(1) :3–23
- White N, McDonald R (2002) Multiple parallel memory systems in the brain of the rat. *Neurobiol Learn Mem* 77 :125–184
- Wiener SI, Taube JS (2005) *Head Direction Cells and the Neural Mechanisms of Spatial Orientation*. MIT Press
- Wilson H, Cowan J (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* 12(1) :1–24
- Wilson S (1991) The animat path to AI. Dans : *Proceedings of the First International Conference on Simulation of Adaptive Behavior*, vol 1, pp 15–21

- de Wit S, Kosaki Y, Balleine B, Dickinson A (2006) Dorsomedial prefrontal cortex resolves response conflict in rats. *J Neurosci* 26(19) :5224–5229
- Yin H, Knowlton B (2004) Contributions of striatal subregions to place and response learning. *Learn Mem* 11(4) :459–463
- Yin H, Ostlund S, Knowlton B, Balleine B (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* 22 :513–523
- Young J, Shapiro M (2009) Double dissociation and hierarchical organization of strategy switches and reversals in the rat PFC. *Behav Neurosci* 123(5) :1028–1035