Université Pierre et Marie Curie
Institut des Systèmes Intelligents et de Robotique

Habilitation à Diriger des Recherches

spécialité « Sciences de la Vie »

par

Benoît Girard

Modélisation neuromimétique : Sélection de l'action, navigation et exécution motrice

Soutenue le 27 septembre 2010 devant le jury composé de :

Dr.	Frédéric Alexandre	INRIA	(Rapporteur)
Prof.	Alain Berthoz	Collège de France/CNRS	(Examinateur)
Prof.	Philippe Bidaud	UPMC/CNRS	(Examinateur)
Prof.	Philippe Gaussier	ENSEA/UCP/CNRS	(Examinateur)
Prof.	Adonis Moschovakis	University of Crete	(Rapporteur)
Dr.	Bruno Poucet	Université de Provence/CNRS	(Rapporteur)

– Donc vous n'avez pas qu'une seule réponse à vos questions ?
– Adso, si tel était le cas, j'enseignerais la théologie à Paris.
– A Paris, ils l'ont toujours, la vraie réponse ?
– Jamais, dit Guillaume, mais ils sont très sûrs de leurs erreurs.
U. Eco, *Le nom de la rose.*

Remerciements

Je voudrais tout d'abord exprimer mes plus profonds remerciements à ceux sans qui l'essentiel des travaux présentés dans ce manuscrit n'existeraient pas ; ceux qui, en tant que stagiaires de M2, doctorants ou post-docs, les ont menés à mes côtés et/ou sous ma direction, les doigts dans le cambouis (ou, plus précisément, dans le clavier) : Francis Colas, Alexandre Coninx, Mariella Dimicolli, Laurent Dollé, Fabien Flacher, Mehdi Khamassi, Sébastien Laithier, Jean Liénard, Cécile Masson, Steve N'Guyen, Quang Cuong Pham, Nicolas Tabareau, Charles Thurat et David Tlalolini-Romero.

Je tiens également à remercier ceux qui, m'ayant accueilli dans leurs laboratoires et équipes, m'ont soutenu et m'ont permis de développer mes problématiques de recherche dans d'excellentes conditions : Agnès Guillot et Jean-Arcady Meyer à l'AnimatLab, Alain Berthoz au LPPA et Philippe Bidaud à l'ISIR.

J'ai également eu la chance de rencontrer, au gré des vents tourbillonnants de la collaboration scientifique multidisciplinaire, des collègues plus expérimentés qui m'ont transmis une part précieuse de leur expérience : Angelo Arleo, Daniel Bennequin, Pierre Bessière, Philippe Gaussier, Laure Rondi-Reig, Jean-Jacques Slotine et Philippe Souères.

Je suis très honoré qu'après avoir interagi avec moi lors de ces nombreuses occasions d'échanges intellectuels qui ponctuent la vie de chercheur (séminaires, colloques, conférences, pots de thèse ou d'HDR, séjours dans un laboratoire), Frédéric Alexandre, Adonis Moschovakis et Bruno Poucet aient accepté de se plonger dans ce manuscrit et d'en être rapporteurs.

Enfin, en septembre 2003, je remerciais Valérie pour son soutien pendant ma thèse et lui souhaitais bon vent pour la sienne. Aujourd'hui je remercie avec reconnaissance et fierté la Dr Durand-Girard pour son soutien pendant la préparation de mon HDR, en particulier durant notre escale forcée à Vancouver, et lui souhaite également bon vent pour son habilitation, lorsque le temps sera venu !

Paris, le 2 septembre 2010.

Table des matières

Introduction

1

Ce manuscrit d'Habilitation à Diriger des Recherches synthétise les travaux de recherche que j'ai effectués depuis ma thèse de doctorat, soutenue le 12 septembre 2003. Ils ont d'abord été menés, de 2003 à 2005, dans le cadre d'un post-doctorat au Laboratoire de Physiologie de la Perception et de l'Action (LPPA, UMR7152), unité mixte CNRS-Collège de France dirigée par le Professeur Alain Berthoz. Après mon recrutement au CNRS en qualité de chargé de recherche (CR2), ils se sont poursuivis, de 2005 à 2008, dans ce même laboratoire, puis, depuis janvier 2009, à l'Institut des Systèmes Intelligents et de Robotique (ISIR, UMR7222), unité mixte CNRS-UPMC dirigée par le Professeur Philippe Bidaud.

1.1 Contexte scientifique

Les capacités de mobilité, d'autonomie, de survie, d'adaptation et de cognition des animaux sont une source d'inspiration persistante en robotique. Ce mouvement de fond, qui prend ses racines dans la robotique fondée sur les comportements (*behavior-based robotics*) de Brooks (1986), a depuis pris différentes formes (approche Animat, biorobotique, neuro-robotique, robotique cognitive, robotique développementale, etc.) chacune correspondant à des nuances dans le degré de biomimétisme, le niveau de description, la finalité, la discipline d'interaction privilégiée (neurosciences, psychologie expérimentale), etc. Ces approches, que je désignerai globalement dans ce document par le vocable de *robotique adaptative*, se sont avérées fructueuses tant pour la mise au point de capteurs et d'effecteurs originaux et efficaces, que pour la conception d'architectures de contrôle cognitives (Meyer et Guillot, 1991, 1994; Guillot et Meyer, 2000, 2001; Webb et Consi, 2001; Meyer et Guillot, 2008).

Par ailleurs, depuis les années 70, les neurosciences comportent un volet de modélisation, couramment dénommé *neurosciences computationnelles* (Sejnowski et al., 1988), ayant pour but de synthétiser, à partir des données expérimentales anatomiques, électrophysiologiques et comportementales, des simulations de diverses régions du cerveau. Ces modèles computationnels ont tout d'abord une valeur explicative : ils proposent des algorithmes décrivant le fonctionnement de circuits neuronaux qui ne peuvent être observés que partiellement dans le cadre des travaux expérimentaux (Gurney, 2009), en une sorte de physiologie synthétique. Ils ont aussi pour objectif de proposer des prédictions concernant le fonctionnement des circuits étudiés, que des études expérimentales complémentaires sont en mesure de tester. Ces modèles ont cependant le défaut d'être désincarnés, or comme le soulignaient déjà Chiel et Beer en 1997, « *The brain*

has a body » –le cerveau a un corps– sous-entendu : on ne peut pleinement comprendre le fonctionnement du cerveau qu'aux commandes d'un corps dont la structure et les propriétés biomécaniques ont évolué concomitamment ; en interaction avec un environnement complexe et dynamique, bien plus complexe et dynamique que ne le sont en général les stimulations imposées aux modèles computationnels désincarnés.

La maturité de l'approche bioinspirée en robotique et le développement de modèles de plus en plus complets de grands circuits cérébraux (rendus possible par l'accumulation de très nombreuses données expérimentales) ont permis la convergence de ces deux axes de recherche vers la *neuro-robotique* (Schaal et al., 2008). Cette discipline s'intéresse spécifiquement à l'étude des modèles neuromimétiques incarnés autonomes, avec le double objectif de (1) permettre des progrès en robotique autonome par l'utilisation d'architectures de contrôle aux capacités d'adaptation proches de celles de l'animal et (2) fournir en retour des systèmes artificiels réels ou simulés, entièrement spécifiés, en interaction avec des environnements dynamiques dans des boucles sensorimotrices fermées, avec lesquels il est alors possible de tester des hypothèses biologiques, de formuler des prédictions et donc de nourrir les neurosciences.

C'est dans ce contexte de la robotique adaptative, des neurosciences computationnelles et de leur interface neuro-robotique que se situent mes travaux de recherche. D'un point de vue méthodologique, j'y conçois et manipule principalement des réseaux de neurones artificiels contraints par les données issues de la neurobiologie expérimentale. La simulation de l'activité de ces réseaux artificiels et sa comparaison avec l'activité observée expérimentalement est l'un des principaux juges de paix de ces modèles. Leur implémentation sur des plate-formes robotiques, si elle apporte son lot de problèmes techniques, est mise en œuvre aussi souvent que possible : d'une part, elle force à la conception de boucles sensorimotrices complètes, évitant par là de trop reporter les problèmes non résolus sur les circuits non modélisés ; d'autre part, elle permet souvent de mettre en évidence des limites ou des propriétés des modèles qui n'apparaissent pas dans des simulations trop simplistes ou trop contrôlées.

1.2 Thèmes de recherche

Les thèmes de recherche que j'aborde concernent principalement trois fonctions fondamentales d'un agent cognitif autonome :

- la sélection de l'action,
- la navigation,
- l'exécution motrice.

Ces thèmes d'apparence divers sont en réalité complémentaires et interagissent dans la continuité, ce que je m'efforcerai de mettre en évidence dans ce manuscrit. Pour chacun d'entre eux, de nombreuses problématiques scientifiques sont ouvertes, parmi lesquelles certaines retiennent particulièrement mon attention, que ce soit dans le cadre des travaux effectués, présentés dans ce manuscrit, ou du programme de recherche qui en découle.

1.2.1 Sélection de l'action

Quels mécanismes permettent de générer la successions d'actions à mettre en œuvre pour assurer la survie, le succès de l'exécution d'une tâche ? Cette question générale, au centre de la sélection de l'action, a été le sujet principal de mes travaux de thèse (Girard et al., 2002, 2003, 2004, 2005a), et est restée au centre de mes préoccupations. Elle sera abordée dans l'ensemble du chapitre 2 :

- Quel est le substrat neural du processus de sélection de l'action ? Comment fonctionne-t-il ? C'est afin de contribuer à l'exploration de ces questions que j'ai participé à l'élaboration d'un nouveau modèle des boucles cortico-basales (Girard et al., 2005b, 2006b, 2008). C'est aussi l'une des questions que nous continuerons d'étudier à l'avenir, dans le cadre du projet EvoNeuro (voir section 5.2.1 en conclusion et Liénard et al. (2010)).
- Quels sont les processus adaptatifs permettant, par l'expérience de l'interaction avec l'environnement, de biaiser les processus de sélection de l'action vers les options les plus profitables ? Nous avons abordé cette question du point de vue de l'apprentissage par renforcement, dans un environnement continu, des variables pertinentes et de leurs combinaisons pour la prise de décision (Khamassi et al., 2004, 2005).
- Quelle motivation, parmi un ensemble de motivations conflictuelles, satisfaire en priorité dans un contexte donné ? Cette question, complémentaire de la précédente, a été effleurée dans (Coninx et al., 2008) et mériterait d'être approfondie, en particulier sous l'angle des bases neurales.
- Quel peut être le rôle éventuel d'une évaluation explicite de l'incertitude dans la prise de décision ? Nous avons abordé cette question très générale dans le cadre d'une tâche de sélection de cible pour les mouvements des yeux, modélisée grâce à la programmation Bayésienne (Colas et al., 2008, 2009).
- Comment les circuits parallèles de sélection dans les ganglions de la base (par exemple oculomoteurs et de navigation) sont-ils coordonnés ? Quel est le substrat neural de cette coordination ? Quel sont les rôles complémentaires des circuits corticaux (et des boucles cortico-basales) vis à vis des circuits sous-corticaux (boucles tecto-basales et cerebello-basales, formation réticulée médiale) ? Comment ces deux niveaux sont-ils coordonnés ? Ces questions, que j'ai encore peu abordées (Girard et al., 2004, 2005a; N'Guyen et al., 2010) et qui sont généralement peu étudiées, constituent une part importante de mon projet de recherche (voir chapitre 5).

1.2.2 Navigation

Les buts ayant été identifiés par la sélection de l'action, comment les atteindre physiquement ? Les animaux ont à leur disposition un vaste répertoire de stratégies de navigation dans lequel ils sont capables de puiser les plus appropriées à un contexte donné. Ce qui peut donc se formuler encore sous la forme d'un problème de sélection, est au centre des thématiques abordées dans le cadre de la navigation, dans le chapitre 3 :

- Quelles stratégies de navigation utiliser afin d'optimiser au mieux diverses contraintes telles que la charge computationnelle, l'énergie dépensée, l'incertitude du résultat ? Une seule stratégie doit-elle être sélectionnée ou la décision de direction de déplacement doit-elle être le résultat d'une combinaison de plusieurs stratégies ? Après avoir exploré cette dernière voie durant ma thèse (Girard et al., 2004, 2005a), j'ai contribué à la proposition d'un modèle d'apprentissage de sélection de stratégies (Dollé et al., 2008, 2010a,b). Ce modèle simule le comportement du rat dans diverses conditions expérimentales testant les choix entre une stratégie d'approche d'indice visuel et une stratégie de planification dans une carte cognitive.
- Quel est le substrat neural de l'intégration de chemin et de sa stratégie associée : le retour au point de départ ? Nous avons proposé un premier modèle du décodage des cellules de grilles du cortex entorhinal permettant la réalisation de cette stratégie sur la base d'un réseau de neurones simple (Masson et Girard, 2009), pour lequel cependant persiste encore le problème d'un paramétrage biologiquement peu plausible. Ce modèle permettant en théorie d'obtenir les coordonnées absolues de l'animal dans un environnement de grande taille, il dépasse le seul cadre du retour au point de départ pour rejoindre celui des stratégies de navigation métriques en général.
- En dehors des stratégies classiques et abondamment modélisées –telles que l'approche d'indice, l'apprentissage de réponses associées à un lieu, ou encore la planification– quelles autres stratégies peuvent être mises en œuvre par les animaux ? Comment fonctionnent-elles ? Quel sont leurs substrats neuraux ? Quelles sont les dynamiques de leurs apprentissages ? Ces questions structurent mes projets portant sur la navigation.

1.2.3 Exécution motrice

La décision d'agir ayant été prise, comment les circuits en charge de la sélection interagissent-ils avec ceux en charge de l'exécution ? Le champ d'étude de la motricité est extrêmement vaste, je me restreins à y aborder quelques sujets d'étude, en lien avec les thèmes ci-dessus, et présentés dans le chapitre 4 :

- Comment encoder et décoder l'espace sensorimoteur pour agir efficacement ? Dans quels référentiels ? C'est dans le cadre des mouvements saccadiques que nous nous sommes spécifiquement intéressés à l'encodage rétinotopique à la surface du colliculus supérieur de la position des cibles dans le champ visuel et de sa transformation en une commande motrice utilisable par les motoneurones extra-oculuaires (Tabareau et al., 2007). Nous avons démontré un lien fondamental entre la géométrie de ces cartes et la manière dont opère cette transformation.
- Dans le cadre de la robotique humanoïde, nous avons commencé à étudier les référentiels (oculo-centrés ou corps-centrés) pour les actions couplées d'atteinte par le regard et le bras (Tran et al., 2009a,b). Cela touche également aux question de la coordination de circuits de

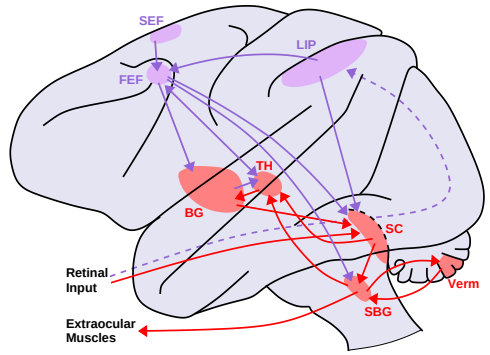

FIG. 1.1 – *Circuits sous-corticaux (rouge) et corticaux (violet) impliqués dans la génération de saccades chez le maccaque. BG : Ganglions de la base, FEF : champs oculaires frontaux, LIP : cortex latéral intra pariétal, SBG : générateurs de saccade du tronc cérébral, SC : colliculus supérieur, SEF : champs oculaires supplémentaires, TH : thalamus, Verm : lobules V et VI du vermis cerebelleux.*

sélection parallèles, ici saccadiques et squeletto-moteurs, évoquées plus haut.

1.3 Substrat neural

Les thèmes de recherche que j'aborde semblent, au vu des connaissances actuelles, fortement liés à un substrat neural composé d'éléments interconnectés : cortex, ganglions de la base, thalamus, colliculus supérieur, cervelet, circuits du tronc cérébral (comme cela est illustré pour le circuit saccadique en Fig. 1.1). Les sections qui suivent présentent donc un certain nombre de données neurobiologiques classiques sur les ganglions de la base, le colliculus supérieur et les régions avec lesquelles ils sont connectés. Elle seront utiles pour la présentation ultérieure des travaux de modélisation réalisés.

1.3.1 Les ganglions de la base

Les ganglions de la base sont fortement impliqués dans la sélection de l'action, tant dans son aspect relevant strictement de la sélection entre actions conflictuelles, que dans celui de l'apprentissage par renforcement des valeurs relatives des différentes actions (voir chapitre 2).

Circuits internes des ganglions de la base

Les ganglions de la base (BG[1], voir Fig. 1.2, gauche et 1.3, gauche) sont un ensemble de noyaux sous-corticaux interconnectés (Redgrave, 2007). Ils sont composés de deux noyaux d'entrée, le striatum d'une part et le noyau subthalamique (STN) d'autre part ; ils possèdent un noyau intrinsèque, le globus pallidus externe (GPe) et deux noyaux de sortie, le globus pallidus interne (GPi) et la substance noire réticulée (SNr). Les noyaux

[1]L'ensemble des acronymes introduits dans ce manuscrit sont récapitulés en fin de document.

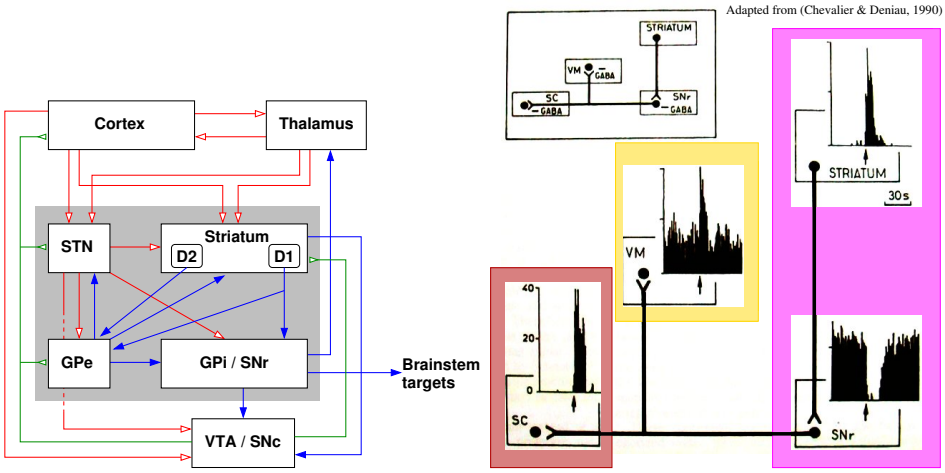

FIG. 1.2 – **Gauche :** *Connectivité des ganglions de la base. D1, D2 : neurones épineux moyens du striatum , ayant respectivement des récepteurs à la dopamine de type D1 ou D2 ; autres abréviations : voir texte. En grisé : noyaux des ganglions de la base ; en rouge : projections glutamatergiques, excitatrices ; en bleu : projections GABAergiques, inhibitrices ; en vert : projections dopaminergiques.* **Droite :** *Desinhibition dans les ganglions de la base. L'activation (ici pharmacologique) du striatum inactive la substance noire réticulée (SNr) qui relâche donc son inhibition sur le colliculus supérieur (SC) et le thalamus ventro-médial (VM), permettant une activité pouvant générer un mouvement saccadique dans le SC et une amplification de l'activité dans VM. Code couleur identique à celui de la Fig. 1.3 :violet : noyaux des ganglions de la base ; brique : colliculus supérieur ; jaune : thalamus. D'après (Chevalier et Deniau, 1990).*

d'entrée reçoivent des entrées glutamatergiques, excitatrices, du cortex (limitées aux aires frontales et préfrontales pour le STN) et du thalamus. Dans les ganglions de la base, seul le STN est également glutamatergique et excitateur, il se projette sur tous les autres noyaux des BG. Les autres noyaux sont inhibiteurs (GABAergiques), le GPe se projette lui aussi sur tous les autres noyaux, alors que le striatum ne cible que le GPe, le GPi et la SNr, et que le GPi et la SNr ne se projettent qu'en sortie vers les cibles des BG :

- le thalamus : noyaux ventro-médian, ventro-antérieur, ventro-latéral, dorsal médian et intra-laminaires,
- des noyaux prémoteurs du tronc cérébral, en particulier le colliculus supérieur (SC) et le noyau pédonculopontin (PPN).

Le striatum, le plus grand noyau des BG, est composé du noyau caudé, du pallidum et du noyau accumbens (NAcc, en position ventrale). En plus des afférences de l'ensemble du cortex, il reçoit, d'une part, dans sa partie ventrale des projections d'origine limbique, en particulier de l'amygdale et de l'hippocampe ; il reçoit également des projections sérotoninergiques du raphé, glutamatergiques et cholinergiques du PPN et dopaminergiques de la substance noire compacte (SNc) et de l'aire ventrale tegmentale (VTA). Le striatum est composé à plus 90% de neurones épineux moyens (MSN), le reste étant constitué d'interneurones de types variés. Parmi ces MSN, ceux comportant des récepteurs à la dopamine de type D1 se projettent vers le GPe, le GPi et la SNr, alors que ceux aux récepteurs de type D2 se projettent exclusivement sur le GPe. La majeure partie des MSN, désignée sous le nom de matrice, est constituée de compartiments neuronaux adja-

cents, les matrisomes, le reste des MSN formant d'autres compartiments isolés au sein de la matrice, les striosomes. Les MSN des matrisomes sont ceux qui se projettent à l'intérieur des ganglions de la base (vers le GPe, le GPi et la SNr), alors que ceux des striosomes semblent se projeter exclusivement sur les noyaux dopaminergiques (VTA et SNc).

L'organisation en compartiments distincts des matrisomes semble se conserver dans l'ensemble des autres noyaux, un compartiment d'un noyau se projetant sur un compartiment associé dans le noyau cible, de sorte que l'on peut distinguer des « canaux » parallèles conservés dans tout le circuit. Ces canaux sont tout de même en mesure d'interagir avec leurs voisins, en particulier via les projections diffuses du STN sur le GPe et le GPi. La structuration en canaux des ganglions de la base est un élément central de l'architecture de tous les modèles computationnels de ce circuit.

Au repos, les sorties des ganglions de la base (GPi et SNr) sont toniquement actives, et maintiennent donc leurs cibles thalamiques et sous-corticales sous inhibition continue. Les BG opèrent par désinhibition (Chevalier et Deniau, 1990) : l'activation d'un canal au niveau du striatum entraîne l'inhibition de ce même canal au niveau des noyaux de sortie, SNr et GPi, ce qui signifie que l'inhibition tonique de ce canal sur ses cibles spécifiques dans le thalamus et le tronc cérébral est levée, permettant une activation ou une amplification de l'activité (Fig. 1.2, droite).

Les ganglions de la base dans le cerveau

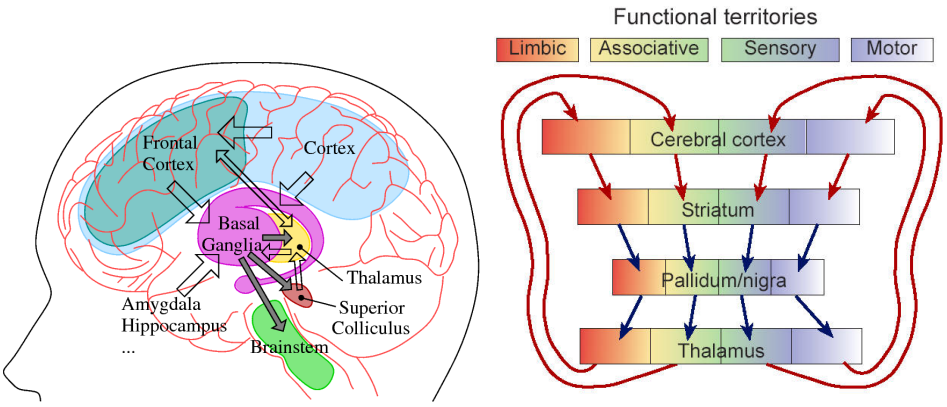

FIG. 1.3 – **Gauche :** *Représentation schématique des interconnexions des ganglions de la base avec le reste du cerveau, en particulier les boucles cortex-ganglions de la base-thalamus-cortex et les boucles colliculus supérieur-thalamus-ganglions de la base-colliculus supérieur. En grisé : projections inhibitrices.* **Droite :** *Représentation schématique des boucles CBTC (reproduit de Redgrave, 2007). Flèches rouges : projections glutamatergiques excitatrices ; flèches bleues : projections GABAergiques, inhibitrices.*

Les BG forment des boucles (Fig. 1.3, gauche) avec des aires corticales (boucle cortico-baso-thalamo-corticales ou CBTC, (Alexander et al., 1986, 1990)) et d'autres structures sous-corticales, par exemple le cervelet (Middleton et Strick, 2000) et le colliculus supérieur (McHaffie et al., 2005). Ces boucles sont dites parallèles car on peut distinguer des sous-circuits disjoints dans les ganglions de la base, chacun ayant la même structure gé-

nérique mais recrutant des sous-parties distinctes des différents noyaux. Ces sous-circuits forment des boucles avec des régions du cortex distinctes (Fig. 1.3, droite) et sont impliquées dans des fonctions distinctes également : on distingue trois grandes catégories (motrices, associatives et limbiques) chacune étant subdivisée en sous-boucles (boucles squelettomotrice et oculomotrice, cette dernière semblant même être subdivisée en une boucle saccadique et une boucle de poursuite lente). Enfin, ces boucles s'empilent suivant un axe ventro-dorsal, le long duquel une structuration hiérarchique semble se dessiner, sur la base de données principalement anatomiques (Joel et Weiner, 1994, 2000; Haber, 2003).

Dopamine

La substance noire compacte (SNc) et l'aire ventrale tegmentale (VTA) ont une action de neuromodulation sur les BG, via des projections dopaminergiques, mais également sur le cortex frontal, l'aire septale, l'amygdale et l'habenula. Les projections dopaminergiques vers les BG ciblent préférentiellement le striatum, mais également le GPe et le STN.

Les neurones dopaminergiques ont une activité tonique ponctuée de bouffées et de creux d'activation. L'activité tonique est supposée liée à l'état motivationnel et à la vigueur des réponses comportementales (Berridge et Robinson, 1998; Kelley, 1999; Niv et al., 2007), suivant le type de récepteur à la dopamine de neurones cible, elle peut avoir un effet excitateur (D1) ou inhibiteur (D2).

La théorie dominante concernant l'activité phasique suppose qu'elle correspond à de l'apprentissage par renforcement au niveau des synapses cortico-striatales (Barto, 1995; Houk et al., 1995). En effet, les bouffées et creux d'activité observés correspondent au calcul de l'erreur de prédiction de récompense utilisé par les algorithmes de type « acteur-critique » (Schultz et al., 1997). Dans ce contexte, le circuit principal des ganglions de la base, passant par les matrisomes du striatum, est supposé jouer le rôle de l'acteur, qui apprend à sélectionner l'action devant rapporter à terme la plus grande récompense. Le circuit passant par les striosomes et les noyaux dopaminergiques serait, lui, le critique en charge d'apprendre à prédire les récompenses futures. Les signaux dopaminergiques phasiques signalent les erreurs de prédiction de récompense et entraînent des modifications des points synaptiques du cortex vers le striatum afin d'à la fois rectifier la prédiction de récompense du critique et modifier le comportement de l'acteur. Se référer à Schultz et al. (1997) pour une description détaillée du mécanisme et à Joel et al. (2002) pour une revue critique de ces théories.

Le rôle des ganglions de la base

Les circuits des ganglions de la base, dont la structure est répétée sans modifications majeures d'une boucle à l'autre, semblent avoir un rôle générique de sélection (Mink, 1996; Redgrave et al., 1999), mis à profit par de nombreuses et diverses fonctions. Cette sélection s'effectue entre les canaux précédemment évoqués. Par exemple, dans le circuit saccadique des BG, chaque canal correspond à un champ moteur spécifique, codant

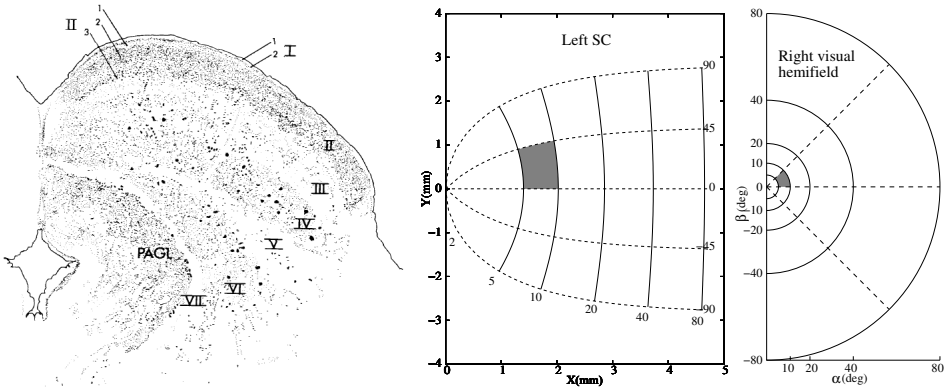

FIG. 1.4 – *Gauche : Couches du colliculus supérieur chez le chat, coupe frontale (reproduit de Kanaseki et Sprague, 1974). Droite : géométrie complexe-logarithmique des carte colliculaires du macaque.*

pour une rotation de l'œil vers une direction donnée dans l'espace (Hikosaka et al., 2000), en compétition pour obtenir le contrôle du prochain mouvement oculaire.

La plasticité des synapses cortico-striatales semble liée à des processus d'apprentissage par renforcement, qu'ils se réduisent aux modèles actuels centrés sur la dopamine ou non. Elle permet de biaiser ce processus de sélection en faveur des choix susceptibles d'être les plus profitables.

1.3.2 Le colliculus supérieur

Le colliculus supérieur est une structure du mésencéphale, composée d'un empilement de sept couches de neurones (Kanaseki et Sprague, 1974, voir Fig. 1.4, gauche), impliqué dans la génération de réponses motrices variées. Plus connu pour son implication dans les mouvements oculaires saccadiques (Moschovakis et al., 1996; Moschovakis, 1996; Scudder et al., 2002), le colliculus est en réalité impliqué de manière plus générale dans les mouvements d'orientation, pouvant recruter les yeux, la tête, voire l'ensemble du corps. La récente démonstration de l'implication du SC dans le cadre de la navigation egocentrée (Felsen et Mainen, 2008) n'est, à ce titre, pas surprenante : un couplage du système d'orientation avec une activité locomotrice est *a priori* suffisant pour une telle navigation. Il a également été montré chez le singe que le colliculus est impliqué dans la génération de mouvements d'atteinte par le bras (Werner, 1993; Werner et al., 1997a,b).

Les neurones du colliculus supérieur

Les couches superficielles du SC contiennent des neurones (V) répondant à des stimulations visuelles (voir Fig. 1.5, A), recevant des entrées projection directes de la rétine pour les plus superficielles, et des entrées issues du cortex visuel pour les autres (Mays et Sparks, 1980). Ces cellules sont organisées en cartes rétinotopiques, chaque colliculus représentant un demi champ visuel. Ces cartes peuvent avoir une géométrie linéaire (chez le rat, la souris, etc.) ou complexe-logarithmique (chez le chat, le singe,

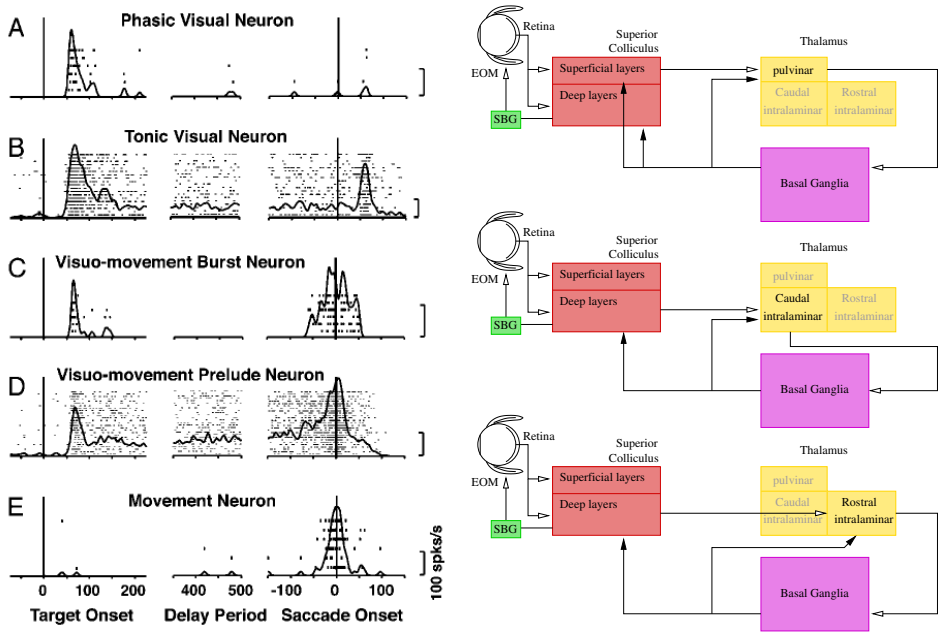

FIG. 1.5 – *Gauche : Catégorisation des cellules du SC selon leur activité, alignées sur la présentation du stimulus (à gauche), durant la période d'attente et alignées sur le déclenchement de la saccade. A : cellule visuelle phasique V (d'autres cellules V ont une activité tonique persistante en présence du stimulus, voir Mays et Sparks (1980)) ; B : cellule quasi-visuelle (QV) ; C : cellules visuo-saccadique (VS) ; D : cellules visuo-saccadiques avec activité maintenue (VMS, probablement les build-up neurons de (Munoz et Wurtz, 1995)) ; E : cellules saccadiques (S). Reproduit de (McPeek et Keller, 2002a). Droite : Boucles tecto-thalamo-baso-tectale, adapaté de McHaffie et al. (2005). EOM : muscles extra-oculaires ; SBG : générateur de saccades.*

voir Fig. 1.4, droite). Les couches plus profondes du colliculus semblent conserver cette organisation particulière.

Plus profondément se trouvent des cellules impliquées dans la mémoire de travail (dites quasi-visuelles ou QV, Mays et Sparks, 1980; McPeek et Keller, 2002a) qui, en plus d'une activité visuelle similaire aux cellules V, sont capables de conserver une activité soutenue après disparition du stimulus visuel (voir Fig. 1.5, B), et de déplacer la localisation de cette activité dans la carte rétinotopique après une saccade afin de conserver une information correcte (*remapping*).

Viennent ensuite les cellules motrices saccadiques, ayant des bouffées d'activité lors de l'exécution d'une saccade (S), et pouvant avoir également une bouffée visuelle (VS) et une activité maintenue entre la présentation du stimulus et la bouffée motrice (VMS, voir Fig. 1.5, C,D et E). Ces cellules motrices se projettent ensuite sur un ensemble de noyaux formant les générateurs de saccade du tronc cérébral (SBG ou Saccade Burst Generators), des circuits coordonnés qui contrôlent les composantes vers le haut, le bas, la gauche et la droite des mouvements saccadiques. Ils se projettent également vers le cervelet (noyaux profonds de la région fastigiale oculomotrice, FOR, en interaction avec les vermis VI et VII) via le noyau reticulé tegmental du pont (NRTP). Enfin, une partie d'entre-eux descend vers la mœlle épinière, permettant par exemple le contrôle des muscles

du cou pour les mouvements d'orientation de grande amplitude mettant à contribution la tête.

C'est encore plus profondément dans le SC que sont localisés les neurones moteurs déchargeant pour des mouvements d'atteinte du bras. Certains d'entre-eux ont une activité purement motrice (R pour *Reach neurons*), alors que d'autres ont aussi une bouffée d'activité visuelle, similaire à celle des neurones visuo-saccadiques (VR), une activité motrice pour les saccades et l'atteinte (SR), voire une combinaison des trois (VSR) (Werner et al., 1997b).

Enfin, dans l'ensemble de ces populations de cellules motrices se trouvent des neurones ayant des réponses multisensorielles (Stein et Meredith, 1993), sensibles non seulement à des stimuli visuels mais également à des stimuli auditifs ou somatosensoriels[2]. Cela soulève d'intéressantes questions de changement de référentiel, la modalité auditive –a priori acquise dans un référentiel cranio-centrique– et la somatosensorielle –dans un référentiel lié au corps– devant donc être recodées dans un référentiel rétinocentrique.

La prise de décision dans le SC et les boucles tecto-basales

Une série de travaux récents a mis en évidence l'implication du SC dans les processus de sélection de cibles des mouvements saccadiques (Basso et Wurtz, 1998; McPeek et Keller, 2002a,b; McPeek et al., 2003; McPeek et Keller, 2004; Carello et Krauzlis, 2004; Krauzlis et al., 2004; Li et Basso, 2005; Li et al., 2006, pour n'en citer que quelques-uns). La plupart d'entre eux supposent cependant que ces processus résultent d'inhibitions latérales dans le SC. L'existence d'au moins trois boucles tecto-thalamo-basotectales clairement identifiées suggère pourtant la possibilité d'une implication des ganglions de la base dans des circuits de sélection purement sous-corticaux (McHaffie et al., 2005, voir Fig. 1.5).

[2]Werner et al. (1997a) signalent l'enregistrement d'un neurone VR ayant également une sensibilité auditive et somatosensorielle.

Sélection de l'Action

2

La sélection de l'action est, pour un agent, le problème du choix à chaque instant de l'action à entreprendre, dans un répertoire en général fini, afin d'atteindre ses objectifs, comme par exemple : assurer sa survie, celle de son espèce ou encore, si l'on considère un agent artificiel, l'exécution d'une tâche prévue par son concepteur. En effet, cet agent peut aussi bien désigner un animal, dans le cadre des sciences de la vie, qu'un robot autonome ou un agent virtuel dans celui de la robotique ou de l'intelligence artificielle : la problématique de la sélection de l'action est typiquement multi-disciplinaire puisqu'étudiée en neurosciences, psychologie, éthologie, intelligence artificielle, sciences politiques, etc. Cela est par exemple illustré par la variété des intervenants au workshop Modeling Natural Action Selection (MNAS) (Bryson et al., 2005) et des auteurs des numéros spéciaux des *Philosophical Transactions of the Royal Society B* (Prescott et al., 2007) et d'*Adaptive Behavior* (Bryson, 2007) qui en découlèrent.

Ainsi qu'évoqué en introduction, un certain nombre de régions du cerveau semble être impliquées dans des processus de sélection de l'action, en particulier les ganglions de la base (voir 1.3.1), dans le cadre de boucles cortico-baso-thalamo-corticales (Mink, 1996; Redgrave et al., 1999; Prescott et al., 1999; Hikosaka et al., 2000; Doya, 2008) et d'autres boucles avec des régions sous-corticales (McHaffie et al., 2005), parmi lesquelles le colliculus supérieur (SC) et le cœur de la formation réticulée médiale (mRF) (Humphries et al., 2005, 2007).

Après des travaux de thèse dédiés à la modélisation des boucles CBTC dans le cadre de la sélection de l'action et de la navigation (Girard et al., 2003, 2005a) sur la base du modèle (GPR) initialement proposé par Gurney et al. (2001a,b), j'ai proposé d'un nouveau modèle des ganglions de la base (CBG) résolvant certaines limitations du GPR et rendant mieux compte de la connectivité interne des BG (2.1) ; j'ai abordé l'apprentissage dans ce type de modèles, en participant à la fusion des modèles de sélection (GPR) avec ceux d'apprentissage par renforcement (acteur-critique), ainsi qu'à l'ajout au CBG d'un module motivationnel adaptatif (2.2) ; enfin, considérant de manière plus abstraite les interactions colliculus supérieur-ganglions de la base, j'ai participé à une étude portant sur la prise en compte explicite de l'incertitude des informations sensorielles dans les processus de sélection, en utilisant le formalisme de la programmation bayesienne et en y intégrant des contraintes neurobiologiques (2.3).

2.1 Modèle contractant des ganglions de la base

Publications : (Girard et al., 2005b, 2006a, 2008)

Dans le cadre de travaux sur la sélection de l'action menés lors de ma thèse, nous avons étudié les propriétés d'un modèle des ganglions de la base proposé par Gurney, Prescott et Redgrave (modèle GPR) (Gurney et al., 2001a,b), évalué dans une tâche de survie minimale (Girard et al., 2002, 2003). Il s'est avéré que le GPR possède d'intéressantes propriétés de persistance dans les choix comportementaux, desquelles pouvaient par exemple découler des économies d'énergie. Ce modèle a cependant laissé entrevoir des problèmes de dynamique interne, se traduisant par exemple par des situations de bloquage durant lesquelles une action précédemment sélectionnée gardait le contrôle de l'agent, malgré l'apparition d'actions compétitrices plus beaucoup pertinentes. Ces problèmes ont été évoqués aussi bien dans Girard et al. (2005a) que dans Prescott et al. (2006). Ce genre d'effet est acceptable s'il a une durée limitée, puisqu'il est à la base des effets bénéfiques de la persistance comportementale, en revanche, s'il perdure trop longtemps, voire indéfiniment, il rend le système totalement inefficace.

Nous nous sommes donc intéressés au développement d'un nouveau modèle des ganglions de la base, dont la dynamique interne serait maîtrisée. Le modèle GPR est composé de neurones à taux de décharge dits « intégrateurs à fuite » qui sont par essence non-linéaires. C'est donc en collaboration avec le Professeur Jean-Jacques Slotine (NSL, MIT) et deux étudiants du LPPA (N. Tabareau et Q.C. Pham) que nous avons utilisé la théorie de la contraction (Lohmiller et Slotine, 1998) afin de concevoir un nouveau modèle des ganglions de la base stable.

2.1.1 Analyse de la contraction et réseaux de neurones

Analyse de la contraction

L'analyse de la contraction permet l'étude de la stabilité exponentielle de systèmes dynamiques non-linéaires de la forme :

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) \tag{2.1}$$

où $\mathbf{x} \in \mathbb{R}^n$, $t \in \mathbb{R}_+$ et $\mathbf{f}$ est une fonction vectorielle $n \times 1$ non-linéaire.

Le principal résultat de l'analyse de la contraction est le théorème suivant (voir Lohmiller et Slotine, 1998, pour la preuve et plus de détails) :

Théorème 2.1 *Considérons le système (2.1) continu dans le temps. Si il existe une matrice uniformément définie positive*

$$\mathbf{M}(\mathbf{x}, t) = \mathbf{\Theta}(\mathbf{x}, t)^T \mathbf{\Theta}(\mathbf{x}, t)$$

telle que le Jacobien généralisé

$$\mathbf{F} = (\dot{\mathbf{\Theta}} + \mathbf{\Theta}\mathbf{J})\mathbf{\Theta}^{-1}$$

est uniformément défini négatif, alors toutes les trajectoires du système convergent exponentiellement vers une unique trajectoire, avec un taux de convergence égal à $|\lambda_{\max}|$, où $\lambda_{\max}$ est la plus grande valeur propre de la partie symétrique de $\mathbf{F}$. Un tel système est dit contractant.

La contraction a de plus l'avantage d'être conservée par assemblage de systèmes contractants dans de nombreuses configurations (hiérarchie, feedback négatif, etc.), moyennant le respect de contraintes portant sur les poids de connexion (Lohmiller et Slotine, 1998; Tabareau et Slotine, 2006).

Neurones à base de systèmes dynamiques localement projetés

La contraction des neurones intégrateurs à fuite posant problème, N. Tabareau et Q.C. Pham ont proposé un nouveau modèle de neurones à taux de décharge fondé sur les systèmes dynamiques localement projetés (Dupuis et Nagurney, 1993; Zhang et Nagurney, 1995, locally Projected Dynamical Systems ou lPDS) sur des $n-$cubes réguliers, fonctionnellement très proches des intégrateurs à fuite, mais dont ils ont prouvé la contraction.

Sur la base d'un opérateur de projection $\mathbf{\Pi}_{\mathbb{H}_n}$ chargé d'assurer le maintien dans les bornes d'un $n-$cube, $\mathbb{H}_n$, dont les bornes représentent les taux de décharge minimum et maximum des n neurones d'un réseau, on peut définir un réseau de neurones de à taux de décharge fondé sur les lPDS :

Définition 2.1 *Un réseau de neurones artificiels de type lPDS est défini par*

$$\dot{\mathbf{x}} = \mathbf{\Pi}_{\mathbb{H}_n}(\mathbf{x}, \mathbf{W}\mathbf{x} + \mathbf{I}(t))$$

où $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))^T$ *est le taux de décharge des neurones,* $\mathbf{W}$ *est la matrice* $n \times n$ *dont la diagonale représente la fuite des neurones et la partie non-diagonale les poids de projections synaptiques entre neurones,* $\mathbf{I}(t)$ *est le vecteur des entrées externes et* $\mathbb{H}_n$ *est un* n-cube régulier.

Or, on peut prouver le théorème suivant pour ce type de réseaux :

Théorème 2.2 *Soit* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$ *un système dynamique contractant dans une métrique constante* $\mathbf{M}$ *compatible avec un ensemble convexe* $\mathbf{\Omega}$*. Alors le lPDS* $\dot{\mathbf{x}} = \mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{f}(\mathbf{x}, t))$ *est aussi contractant dans la même métrique et avec le même taux de contraction.*

Les détails de la définition de $\mathbf{\Pi}_{\mathbb{H}_n}$ et la preuve de ce théorème sont explicités dans l'article (Girard et al., 2008), inclus en section 6.1.

2.1.2 Modèle des ganglions de la base

Le modèle contractant des boucles CBTC est donc un réseau de neurones artificiels de type « lPDS projeté sur un $n-$cube régulier ». Sa structure est représentée sur la Fig. 2.1 et peut être décrite comme suit : chaque action élémentaire en compétition se voit assigner un canal (voir section 1.3.1), représenté dans chaque noyau des ganglions de la base, chaque noyau du thalamus et chaque aire corticale modélisée, par un lPDS (lui-même modélisant l'activité d'un groupe de neurones réels). Quelques modules où la ségrégation en canaux ne semble pas conservée (interneurones rapides du striatum FS et noyau thalamique réticulé TRN, voir plus bas) font exception, ils sont alors modélisés par un unique lPDS.

L'entrée du modèle est un ensemble de *saillances* S, qui sont supposées résulter de la combinaison d'informations sensori-motrices en provenance du cortex, représentant pour chaque action élémentaire la pertinence de

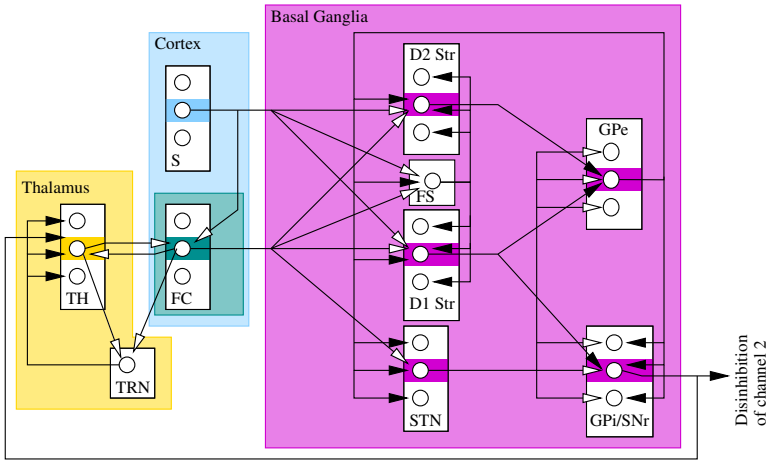

FIG. 2.1 – *Structure du modèle contractant des CBTC. Exemple comportant trois canaux, où le second canal est mis en évidence par les bandes de couleur et où seules ses connexions sont représentées pour plus de clarté. Code couleur identique à celui de la Fig. 1.3. FCtx : cortex frontal, FS : interneurones Fast Spiking du striatum, GPe : globus pallidus externe, GPi : globus pallidus interne, S : saillances (cortex pariétal), SNr : substance noire réticulée, STN : noyau subthalamique, TH : noyau du thalamus, TRN : noyau réticulé du thalamus.*

l'activation de cette action. Les ganglions de la base vont s'efforcer de ne désinhiber que les canaux correspondant aux saillances les plus fortes. La boucle auto-excitatrice entre les noyaux thalamiques et le cortex frontal tend à amplifier les saillances, mais l'influence inhibitrice de la sortie des ganglions de la base ne doit permettre cette amplification que sur les canaux sélectionnés.

Le processus de sélection dans le module BG du modèle s'appuie sur plusieurs propriétés du circuit :

1. Seuil d'activation des MSN : les propriétés d'état bas ou haut des MSN sont modélisées de manière simpliste par une entrée constante négative, qui filtre les saillances inférieures à ce seuil.
2. Inhibition feed-forward : une partie des interneurones du striatum, dits GABAergiques rapides (Tepper et Bolam, 2004; Tepper et al., 2004, fast spiking neurons ou FS), sont inclus dans le modèle, représentés par un seul lPDS du fait de leur couplage supposé. Ils somment l'ensemble des saillances et redistribuent une inhibition proportionnelle à ce signal dans le striatum, et participent donc aussi au filtrage des saillances faibles.
3. Réseaux « off-center on-surround » : les projections inhibitrices ciblées des neurones D1 du striatum et excitatrices diffuses du STN sur le GPi/SNr (et symétriquement pour D2 et STN sur GPe) permettent à chaque canal d'exciter la sortie de ses voisins et d'inhiber sa propre sortie. Dans cette configuration, le canal le plus activé en entrée est susceptible d'exciter ses concurrents plus qu'ils ne s'auto-inhibent, tout en s'inhibant assez fort lui-même pour contrebalancer les excitations de ses concurrents.
4. Boucles auto-inhibitrices : les projections canal à canal du GPe sur le striatum (D1 et D2) permettent de renforcer le contraste de la

sélection : par défaut, le GPe est toniquement actif et inhibe donc l'ensemble du striatum, mais si un canal parvient, via le mécanisme « off-center on-surround », à se sélectionner, et donc à désactiver son canal dans le GPe, alors l'inhibition sur son entrée par ce même GPe baisse, ce qui renforce encore sa sélection.

Les inhibitions feedforward et les boucles auto-inhibitrices sont des ajouts vis-à-vis du GPR. Les inhibitions feedforward ne sont cependant pas une nouveauté, un tel mécanisme ayant déjà été intégré au modèle de Beiser et Houk (1998). Les projection du GPe vers le striatum n'ont, en revanche, encore jamais été intégrées à un modèle computationnel, bien qu'elles aient été documentées depuis longtemps (Staines et al., 1981; Bevan et al., 1998; Kita et al., 1999). Les résultats de Bevan et al. (1998) indiquent avec certitude que ces projections ciblent les interneurones inhibiteurs, ce qui dans notre modèle permet de réguler le niveau d'activation des FS, ils n'excluent cependant pas des projections ciblées vers les MSN, qui se sont avérées particulièrement efficaces dans notre modèle pour assurer une bonne sélection. Enfin, toujours en comparaison avec le GPR, les projections du GPe sur le GPi sont ici diffuses. En effet, les inhibitions canal à canal du GPR *détériorent* la sélection et n'ont aucune fonctionnalité, ce qui explique pourquoi les poids de cette projection étaient à un niveau relativement bas. Les différences de nature des projections du GPe sur le striatum, d'une part, et sur le STN, le GPi et la SNr, d'autre part (Sato et al., 2000; Parent et al., 2000), semblent pouvoir justifier notre choix de projections ciblées dans le premier cas et diffuses dans l'autre.

Au final, en utilisant la formulation proposée par la définition 2.1, le modèle peut s'écrire sous la forme suivante :

$$\begin{aligned}
(\mathbf{Wx}+\mathbf{I}(t))_{D1_i} &= \tfrac{1}{\tau}\big((1+\gamma)(w^{D1}_{FC}x^{FC}_i - w^{D1}_{GPe}x^{GPe}_i + w^{D1}_{S}S_i(t)) \\
&\quad -w^{D1}_{FS}x^{FS} + I_{D1}\big) \\
(\mathbf{Wx}+\mathbf{I}(t))_{D2_i} &= \tfrac{1}{\tau}\big((1-\gamma)(w^{D2}_{FC}x^{FC}_i - w^{D2}_{GPe}x^{GPe}_i + w^{D2}_{S}S_i(t)) \\
&\quad -w^{D2}_{FS}x^{FS} + I_{D2}\big) \\
(\mathbf{Wx}+\mathbf{I}(t))_{FS} &= \tfrac{1}{\tau_{FS}}\textstyle\sum_{j=1}^{N}\Big(w^{FS}_{FC}x^{FC}_j - w^{FS}_{GPe}x^{GPe}_j + w^{FS}_{S}S_j(t)\Big)
\end{aligned} \tag{2.2}$$

$$\begin{aligned}
(\mathbf{Wx}+\mathbf{I}(t))_{STN_i} &= \tfrac{1}{\tau_{STN}}\Big(w^{STN}_{FC}x^{FC}_i - w^{STN}_{GPe}\textstyle\sum_{j=1}^{N}x^{GPe}_j + I_{STN}\Big) \\
(\mathbf{Wx}+\mathbf{I}(t))_{GPe_i} &= \tfrac{1}{\tau}\Big(-w^{GPe}_{D1}x^{D1}_i - w^{GPe}_{D2}x^{D2}_i + w^{GPe}_{STN}\textstyle\sum_{j=1}^{N}x^{STN}_j + I_{GPe}\Big) \\
(\mathbf{Wx}+\mathbf{I}(t))_{GPi_i} &= \tfrac{1}{\tau}\Big(-w^{GPi}_{D1}x^{D1}_i + w^{GPi}_{STN}\textstyle\sum_{j=1}^{N}x^{STN}_i \\
&\quad -w^{GPi}_{GPe}\textstyle\sum_{j=1}^{N}x^{GPe}_j + I_{GPi}\Big) \\
(\mathbf{Wx}+\mathbf{I}(t))_{TH_i} &= \tfrac{1}{\tau_{TH}}\big(w^{TH}_{FC}x^{FC}_i - w^{TH}_{TRN}x^{TRN} - w^{TH}_{GPi}x^{GPi}_i\big) \\
(\mathbf{Wx}+\mathbf{I}(t))_{FC_i} &= \tfrac{1}{\tau_{FC}}\big(w^{FC}_{S}S_i + w^{FC}_{TH}x^{TH}_i\big) \\
(\mathbf{Wx}+\mathbf{I}(t))_{TRN} &= \tfrac{1}{\tau_{TRN}}\big(\textstyle\sum_i w^{TRN}_{FC}x^{FC}_i + w^{TRN}_{TH}x^{TH}_i\big)
\end{aligned}$$

où N est le nombre de canaux, $i \in [1, N]$ et γ est le niveau de dopamine tonique. Les valeurs numériques des paramètres utilisés sont précisées dans l'article (Girard et al., 2008), inclus en section 6.1

2.1.3 Résultats

Contraction du modèle

L'analyse de la contraction du module ganglions de la base, d'une part, et du module thalamo-cortical, d'autre part, a permis d'identifier les conditions suivantes sur les poids des connexions pour que chacun de ces modules soit contractant :

$$\begin{cases} ((1+\gamma)w_{D1}^{GPe}w_{GPe}^{D1})^2 + ((1-\gamma)w_{D2}^{GPe}w_{GPe}^{D2})^2 & < & 1 \\ w_{TH}^{FC}(w_{FC}^{TH} + \sqrt{{w_{FC}^{TH}}^2 + N{w_{FC}^{TRN}}^2}) & < & 1 \end{cases} \tag{2.3}$$

La complexité de la Jacobienne généralisée du système complet (modules ganglions de la bas et boucle thalamo-corticale connectées) rend difficile le calcul d'une solution algébrique globale sur les poids de projection assurant qu'elle reste définie négative. Il a en revanche été vérifié, en calculant numériquement les valeurs propres de la partie symétrique de la Jacobienne généralisée pour les paramètres fixés, qu'elle sont bien toutes négatives.

Propriétés de sélection désincarnées

Le respect des conditions de contraction assure que, quel que soit l'état interne du modèle à un instant donné, si l'on fixe les saillances en entrée, le système va converger exponentiellement vite vers un unique état, sans rien présumer de la nature de cet état. Une première série de paramètres du modèle ont été choisis de sorte que, au repos (saillances nulles), le STN, le GPe et le GPi/SNr aient une activité non-nulle, afin de respecter les données électrophysiologiques. Les paramètres restants ont ensuite été ajustés pour que, dans le cas de saillances non-nulles, le canal ayant l'entrée la plus forte soit désinhibé le plus parfaitement possible (sortie de GPi/SNr la plus proche de 0 possible), et pour que les autres canaux soient inhibés, c'est-à-dire que l'activité de GPi/SNr soit supérieure ou égale à son activité au repos.

Nous avons ensuite reproduit le test proposé par Gurney et al. (2001b) dans leur article décrivant le GPR avec la dernière version du GPR présentée dans Prescott et al. (2006) et avec le CBG (voir Fig. 2.2). Ce test met en évidence le fait que le GPR n'est pas contractant, puisqu'à l'étape 4, alors que les deux canaux 1 et 2 ont la même saillance, seul le canal 2 est sélectionné. Ceci, car il avait la main à l'étape précédente : *le GPR converge vers un état différent selon ses conditions initiales*.

Nous avons également reproduit l'exploration de saillances systématique sur deux canaux proposée dans Prescott et al. (2006). Sans entrer dans les détails, ce calcul confirme que le GPR n'est pas contractant, puisque sa sélection exhibe un comportement d'hystérésis lorsque l'on fait varier progressivement des saillance de deux canaux en compétition.

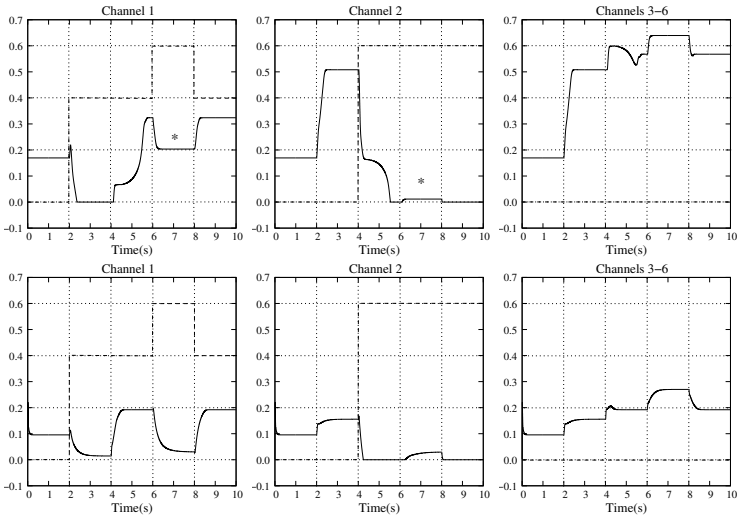

FIG. 2.2 – *Évolution des sorties inhibitrices du GPi/SNr dans le test en 5 étapes proposé par Gurney et al. (2001b), pour (en haut) le GPR et (en bas) le CBG. Le ou les canaux ayant les saillances les plus élevées sont supposés être moins inhibés que la valeur de référence au repos, obtenue à l'étape 1. Les lignes en pointillé représentent les saillances d'entrée. Durant la 4ème étape, ($6s < t < 8s$), les canaux 1 et 2 sont bien sélectionnés par le CBG alors que le GPR ne sélectionne que le 2 (astérisque).*

Tâche simulée de survie

Une telle dépendance aux conditions initiales est-elle souhaitable, dans l'absolu, pour un mécanisme de sélection de l'action ? Aucune réponse définitive n'a été proposée à cette question.

Dans leur article de 2006, Prescott et al. prétendent que cette hystérésis permet d'éviter des oscillations comportementales dans le cadre d'une implémentation robotique. Ces oscillations apparaissent lorsque deux actions en compétition ont des saillances très proches, et lorsque l'activation d'une action fait baisser sa propre saillance (c'est couramment le cas, par exemple avec des comportements de recharge), ce qui fait que l'action concurrente prend la main, et ainsi de suite. Les test du GPR que nous avions effectués avec une tâche de survie (Girard et al., 2003) montraient effectivement que le GPR pouvait éviter de telles oscillations néfastes.

Pour autant, le CBG peut avoir une dynamique de convergence assez lente lorsque deux actions ont des saillances proches (voir par exemple l'étape 4 du test de la Fig. 2.2), tout en évitant les blocages (la convergence vers un unique état étant assurée par contraction). Ceci pourrait être tout à fait suffisant pour résoudre les problèmes d'oscillations comportementales. Nous l'avons donc testé dans une tâche de survie simulée similaire à celle proposée dans Girard et al. (2003), afin de savoir si une hystérésis était nécessaire pour éviter les oscillations, et avons comparé ses performances à un enchaînement de règles *if-then-else* (ITE) sans mémoire.

Dans cette tâche de survie, le robot est doté d'un métabolisme simulé comportant deux variables internes : l'énergie E et l'énergie potentielle E_p, variant entre 0 et 1. Le robot consomme en permanence de l'énergie (10^{-2} unités par seconde), et son seul moyen d'en récupérer est de recharger de l'E_p sur des sources d'E_p disséminées dans l'environnement, puis de

la transformer en E sur d'autres sources dédiées (voir Girard et al., 2003, pour plus de détails). Dans l'implémentation de ce problème que nous avons testée, il doit effectuer un choix entre 7 actions élémentaires :

1. Exploration aléatoire (W).
2. Évitement d'obstacle (AO).
3. Approche d'une source d'E (AE) : si une source d'E est visible par la caméra, le robot s'oriente vers cette source tout en avançant.
4. Approche d'une source d'E_p (AE_p).
5. Recharge E (ROE) : transforme de l'E_p en E si une source d'E est assez proche et s'il y a de l'E_p en stock.
6. Recharge E_p (ROE_p) si une source d'E_p est assez proche.
7. Repos (Sl) : consomme moitié moins d'E que les autres, mais le robot est immobile et ne peut donc rechercher les sources dont il a besoin, ne peut donc être utilisé sans risque que lorsque les stocks d'E et d'E_p sont élevés.

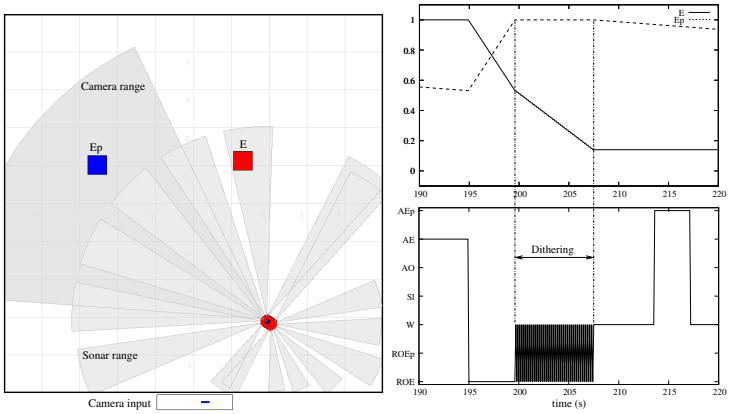

FIG. 2.3 – *Tâche de survie.* ***Gauche :*** *environnement de test de* $10 \times 10m$*, carré bleu : source d'*E_p*, carré rouge : source d'*E *, gris clair : sonars, gris foncé : angle de vue et portée de la caméra.* ***Droite :*** *exemple typique d'oscillations comportementales entre la recharge d'énergie et l'exploration, en haut : niveaux d'*E *et d'*E_p*, en bas : action sélectionnée. Noter que pendant les oscillations,* 0.3 *unités d'*E_p *on été consommées en* $7s$ *alors qu'elles devraient permettre d'en survivre* 30*.*

Le détail des entrées sensorielles fournies par le simulateur, des calculs des saillances pour ces sept actions pour le CBG, et la règle ITE de comparaison, sont fourni dans les annexes de l'article, en section 6.1).

Alors que la charge initiale d'énergie ne permet de survivre que $100s$, le modèle CBG aussi bien que la règle ITE sont capables de survivre en moyenne (sur 20 essais), $687s$ ($\sigma = 244$) et $737s$ ($\sigma = 218$) respectivement, et les deux ensembles de durées de survie ne proviennent pas de distributions significativement différentes (test de Kolmogorov-Smirnoff : $p = 0.771$). Cependant, les comportements des deux systèmes pour obtenir ces performances similaires sont différents : la règle ITE, ne possédant pas de mémoire, rencontre des situations où elle ne peut qu'osciller. En effet, lorsque le robot termine de se recharger en énergie, il entame un déplacement qui ne l'éloigne pas beaucoup de la source, et comme un peu

d'énergie a été consommée par ce déplacement, il s'arrête à nouveau pour se recharger, souvent jusqu'à ce qu'il n'ait plus d'énergie potentielle en stock (Fig. 2.3, droite). Les stocks d'E_p étant donc souvent bas, la règle ITE n'active pas souvent le comportement de repos qui permet d'économiser l'énergie.

Ces oscillations dissipent de surcroît de l'énergie inutilement : en un pas de temps d'exploration, le robot dépense 10^{-2} unités d'E, et en un pas de temps de recharge il consomme 0.2 unités d'E_p, qui devraient générer 0.2 unités d'E, mais la réserve d'E ayant un maximum de 1, seules les 10^{-2} unités consommées sont récupérées, le reste est perdu. Au final, l'énergie potentielle moyenne extraite de l'environnement par chacun des système pour survivre $1s$ est significativement différente (test KS : $p < 0.001$) : le CBG, qui exploite le comportement de repos, consomme $0.93 \times 10^{-2} Ep.s^{-1}$ ($\sigma = 0.30 \times 10^{-3}$), alors que la règle ITE, avec ses oscillations, consomme $1.17 \times 10^{-2} Ep.s^{-1}$, ($\sigma = 1.17 \times 10^{-3}$).

En conclusion, la tâche proposée est susceptible de donner lieu à des oscillations comportementales coûteuses, et le CBG est paramétrable de manière à exhiber une persistance dans le choix de ses actions permettant d'éviter ce problème. Le phénomène d'hystérésis dans la sélection des actions du GPR n'est donc pas indispensable pour faire face aux oscillations comportementales, alors que la contraction établie du CBG permet de maîtriser sa dynamique.

2.1.4 Discussion

Le travail réalisé dans cette étude a permis la proposition d'un nouveau modèle des processus de sélection dans les ganglions de la base intégrant des connexions négligées par les précédents modèles et renforçant l'efficacité de la sélection. Pour autant, d'autres connexions documentées dans la littérature expérimentale sont encore laissées de côté, comme par exemple celles d'un sous-groupe de neurones du STN projetant vers de striatum. Cette projection n'est pas intégrée dans les modèles actuels par manque de données physiologiques permettant de cerner son rôle, mais aussi par manque de propositions théoriques quant à l'utilité d'une telle projection. Il en va de même pour la modulation dopaminergique qui, au delà du striatum, affecte le STN et le GPe. C'est précisément pour favoriser l'exploration de nouvelles propositions de modèles computationnels, ici nécessaire, que la conception automatique de modèles via l'usage d'algorithmes évolutionnistes contraints par les connaissances neurobiologiques a été proposée dans le projet collaboratif EvoNeuro (voir section 5.2).

Bien que non-linéaire, la dynamique du modèle CBG est maîtrisée par l'utilisation de l'analyse de la contraction qui en garantit la convergence exponentielle. Cela permet d'éviter les problèmes de blocage de la sélection, préalablement rencontrés avec le modèle GPR. Les propriétés de préservation de la contraction par combinaison de systèmes contractants se sont avérées particulièrement utiles dans ce cadre, la contraction ayant d'abord été étudiée pour les ganglions de la base, puis pour la boucle thalamo-corticale, avant d'étudier celle du circuit complet. On peut espérer étendre cette approche à des circuits neuronaux plus grands encore, comme par exemple l'ensemble du système saccadique (voir section 5.1.3).

Un sous-produit de cette approche théorique aura été la proposition d'un nouveau modèle de neurones à taux de décharge, sur la base de « systèmes dynamiques projetés localement sur des hypercubes réguliers », dont la contraction est prouvée. L'intérêt de l'usage généralisé de ce modèle en lieu et place des traditionnels « intégrateurs à fuite » mériterait d'être étudié.

En ce qui concerne le problème général de la sélection de l'action, ce modèle aura permis de montrer que la résolution d'une tâche de survie simple à deux sources ne nécessite pas d'effet d'hystérésis dans le système de sélection, contrairement à ce qu'avançaient Prescott et al. (2006). Cela ne plaide pas en la faveur de l'inclusion de tels comportements dans la conception de systèmes de sélection de l'action, qu'ils soient neuromimétiques ou non.

Enfin, les calculs de saillances choisis pour la résolution d'une tâche (détaillés en dans l'annexe B de l'article fourni en section 6.1 pour notre tâche en particulier), sont le résultat d'un processus de paramétrage relativement fastidieux et dont il n'est guère assuré qu'il ait convergé vers un optimum. Les ganglions de la base étant un substrat neural fondamental pour l'apprentissage par renforcement, une extension naturelle des modèles des processus de sélection des ganglions de la base est l'adjonction de processus d'apprentissage par renforcement.

2.2 Apprentissage et Adaptation de la sélection de l'action

Les implémentations robotiques, réelles ou simulées, des modèles des ganglions de la base utilisés comme mécanismes de sélection de l'action (Girard et al., 2003, 2005a; Prescott et al., 2006; Girard et al., 2008) ne sont efficaces que si un ajustement minutieux des saillances d'entrée est effectué par le modélisateur. Cet ajustement nécessite à la fois de trouver les bonnes combinaisons de variables et leur juste pondération. Nous avons montré dans Girard et al. (2003) que même dans le cas simple d'une tâche de survie minimale, les combinaisons de variables nécessaires à une sélection efficace ne peuvent se limiter à des sommes pondérées, mais doivent intégrer des entrées de type *sigma-pi* (combinant sommes et multiplications), ainsi que des fonctions de transfert non-linéaires. Concernant les pondérations, elles servent principalement à définir les priorités relatives entres actions élémentaires, ainsi, lorsque le robot est sur sa station de recharge et manque d'énergie, l'action de recharge doit être prioritaire sur l'action d'approche de la station.

La conception d'agents capables de s'adapter à des environnements changeants nécessite que les ajustements des calculs de saillances puissent être modifiés en ligne, de manière autonome pendant toute la durée de vie de l'agent. Du point de vue des modèles des BG, ce paramétrage des entrées du système s'opère au niveau des synapses cortico-striatales, qui sont précisément un lieu de plasticité neurale sous contrôle des noyaux dopaminergiques mésencéphaliques (voir section 1.3.1).

J'ai donc participé à deux études sur ce sujet, la première portait sur l'adaptation des modèles standards d'apprentissage par renforcement

acteurs-critiques pour l'ajustement des saillances d'un modèle GPR (Khamassi et al., 2004, 2005, co-encadrement du stage de M2 de M. Khamassi) ; la seconde s'est intéressée à la modulation relative, par un système motivationnel, de la priorité de groupes d'actions élémentaires associés à des sources différentes de l'environnement tout en conservant les priorités entre actions dans ces groupes (Coninx et al., 2008, encadrement du stage de M2 d'A. Coninx).

2.2.1 Modèles acteurs-critiques

Publications : (Khamassi et al., 2004, 2005)

La similarité des profils d'activation des neurones dopaminergiques avec les prédictions des modèles « acteur-critique » d'apprentissage par renforcement a été soulignée en section 1.3.1. Cependant, les modèles proposés considèrent en général des systèmes de sélection (composante acteur) extrêmement simplifiés, très éloignés de la complexité structurelle des ganglions de la base. De plus, ils utilisent des critiques composés d'un seul ou de quelques neurones artificiels, dont les capacités de discrimination sont limitées. Cela ne pose pas de problème pour les tâches simplifiées, aux états discrets connus à l'avance, dans lesquelles ils ont été testés. Nous avons donc cherché à étudier l'impact de l'utilisation d'un modèle neuromimétique du circuit acteur (en l'occurrence le modèle GPR) et de quatre configurations différentes du circuit critique, dans l'apprentissage d'une tâche de rechargement dans un environnement simulé continu et avec des entrées sensorielles nombreuses.

Modèle

Les « acteurs » utilisés dans les études de modélisation de l'apprentissage par renforcement dans les ganglions de la base sont de simples processus « winner-takes-all » (WTA) : pour chaque action élémentaire, une somme pondérée des entrées est calculée, et c'est l'action pour laquelle cette saillance est maximale qui est choisie. C'est une abstraction assez poussée de la complexité du réseau des ganglions de la base décrit en section1.3.1. Les modèles des processus de sélection dans les ganglions de la base ont un fonctionnement qualitatif similaire à un WTA, cependant, ils ont une dynamique interne résultant en une certaine inertie des sélections effectuées qui peut avoir des effets bénéfiques sur l'ensemble du comportement (voir la section précédente, 2.1). Il semblait donc intéressant d'en intégrer un, en l'occurrence le GPR (Gurney et al., 2001b), en lieu et place d'un WTA.

Le modèle de « critique » initialement proposé par Houk et al. (1995) est un unique neurone, aux capacités calculatoires limitées. Dans le cadre de résolution de tâches d'apprentissage complexes, d'autres propositions ont été faites, elles sont cependant contraintes par l'hypothèse que le critique est localisé dans les striosomes du striatum. Elles sont donc limitées à des réseaux à une couche, fonctionnant sur le principe des mixtures d'experts (Jacobs et al., 1991), où des réseaux simples se partagent la tâche d'apprentissage en se spécialisant chacun sur une sous-région de l'espace

d'état. Nous avons cherché à comparer les mérites de quatre de ces différentes architectures soumises à une même tâche :

1. AC : Un acteur GPR connecté à un critique composé d'un unique neurone, similaire à la proposition de Houk et al. (1995).
2. AMC1 : Un acteur GPR connecté à N critiques, ces critiques sont en compétition (celui qui a le mieux prédit apprend le plus), utilisant un algorithme similaire à celui de Jacobs et al. (1991).
3. AMC2 : Un acteur GPR connecté à N critiques, dont la répartition dans l'espace d'état est fournie a priori (de manière similaire à Suri et Schultz, 1998)
4. MAMC : N acteurs GPR, chacun connecté à un critique, la répartition des couples acteurs-critiques dans l'espace d'état est donnée a priori (approche similaire à celles de Baldassarre, 2002; Doya et al., 2002).

Dans toutes ces propositions, $N = 30$.

Tâche

Ces modèles on été testés dans une tâche de labyrinthe en croix, simulant un protocole utilisé dans des expériences de navigation chez le rat (Albertin et al., 2000). Il s'agit de la phase préliminaire de l'entraînement, où les rats apprennent qu'une lumière au bout d'un des bras du labyrinthe (Fig. 2.4, gauche) est associée à une récompense : un bras sélectionné aléatoirement voit son extrémité illuminée, si le rat s'en approche, il reçoit une récompense (deux gouttes d'eau), le bras est éteint et un autre bras est allumé.

Le robot simulé est placé dans un environnement similaire, de $5m$ de large. Les murs sont noirs, l'extrémité des bras est soit gris-foncé (éteinte), soit blanche (allumée). Le centre du labyrinthe est marqué d'une croix gris-clair permettant de l'identifier visuellement.

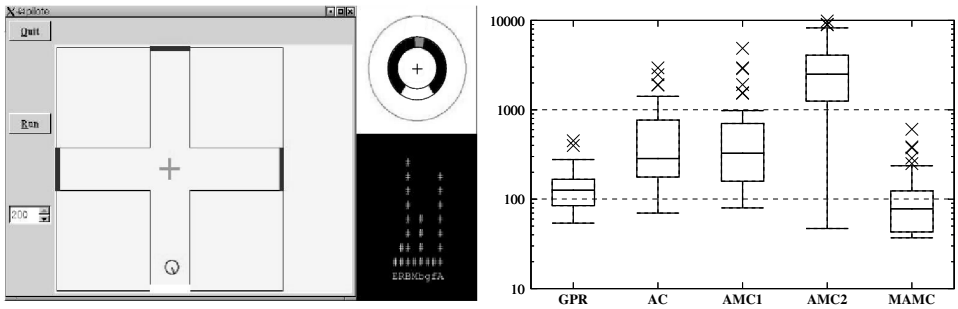

FIG. 2.4 – **Gauche :** *Tâche du labyrinthe en croix. Depuis chaque bras sombre, le robot doit apprendre à rejoindre le centre, à se diriger vers le bras éclairé, et à se recharger lorsqu'il est au bout. En haut à droite, les perceptions visuelles du robot, en bas à droite, histogramme des saillances des différentes actions.* **Droite :** *Statistiques des durées des 50 derniers essais pour chaque algorithme testé (échelle logarithmique). Barre centrale : médiane, boîte : premier (Q_1) et troisième (Q_3) quartiles, moustaches : extremums, croix : données aberrantes (n'appartiennent pas à l'intervalle $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$.).*

Le robot est équipé d'une caméra linéaire monochrome, d'une définition de 10 degré par pixel (Fig. 2.4, gauche). A partir de cette image

sont calculées quatre variables pour chaque couleur (blanc, gris-clair, gris foncé) : la première indique si la couleur est visible, et si elle l'est, la seconde fournit l'angle que fait la tache de couleur avec la direction de déplacement actuelle, la troisième est sa largeur en pixels et la dernière sa distance approximative. Avec une entrée constante à 1, le modèle a donc 13 variables d'entrée.

Le robot doit choisir parmi les actions élémentaires suivantes :

1. Avancer (A) tout droit à la vitesse de $40cm.s^{-1}$.
2. S'orienter vers le Blanc (b) à la vitesse de $10^{\circ}.s^{-1}$.
3. S'orienter vers le Gris clair (g) à la vitesse de $10^{\circ}.s^{-1}$.
4. S'orienter vers le Gris foncé (f) à la vitesse de $10^{\circ}.s^{-1}$.
5. Boire (B), efficace si le robot est à moins de $30cm$ face au mur blanc, il reçoit alors successivement deux récompenses ($R = 1$).
6. Repos (R), c'est-à-dire rester immobile.

Un certain nombre d'entrées sensorielles (informations sur les objets gris-foncé) et d'actions élémentaires (orientation vers le gris-foncé, repos) ne sont pas utiles à la résolution de la tâche. L'espace sensorimoteur dans lequel l'apprentissage est effectué n'est donc pas a priori restreint aux seuls éléments pertinents, ce qui est le cas général de l'apprentissage. Le processus d'apprentissage, pour résoudre cette tâche, devra donc être capable d'identifier ce dont il ne doit pas tenir compte ou ne pas faire usage.

Résultats

Le temps nécessaire à l'obtention d'une récompense, qui se doit d'être le plus faible possible, est la mesure de performance utilisée. Après un entraînement de 50 essais, les durées des 50 essais suivants sont mesurées pour chaque modèle (Fig. 2.5) ainsi que pour un modèle GPR dont les calculs de saillances ont été ajustés à la main. Leurs statistiques sont comparées (Fig. 2.4, droite), en particulier leurs médianes (test de Wilcoxon) et leurs distributions (test de Kolmogorov-Smirnoff).

Le modèle AC classique n'améliore plus sa performance bien avant la fin des 50 essais d'entraînement (Fig.2.5), et reste à des durées plus élevées que le GPR ajusté à la main (test de Wicoxon : $p < 0.001$, Fig. 2.4, droite). Cela vient des capacités limitées de son critique : il ne parvient à apprendre à prédire correctement les récompenses qu'à proximité du but (dans le bras illuminé), et fait des choix aléatoires dans le reste du labyrinthe.

Le modèle AMC1 est supposé compenser les limitations du modèle AC en utilisant plusieurs critiques. Ils sont contrôlés par un « gating network » (Jacobs et al., 1991) en charge de faire apprendre prioritairement le critique qui propose les meilleures prédictions, pour aboutir à une spécialisation de chaque critique dans un sous domaine de l'espace d'état. En l'occurrence, cette spécialisation échoue : un seul expert prend la main dans tout le labyrinthe. Les performances sont alors similaires à celles du modèle AC : pas de différence significative des médianes des durées (test de Wilcoxon, $p = 0.51$) ni de leurs distributions (test KS, $p = 1$).

Le modèle AMC2 cherche à contourner le problème rencontré avec l'AMC1, en prédéfinissant les sous-régions associées à chaque critique.

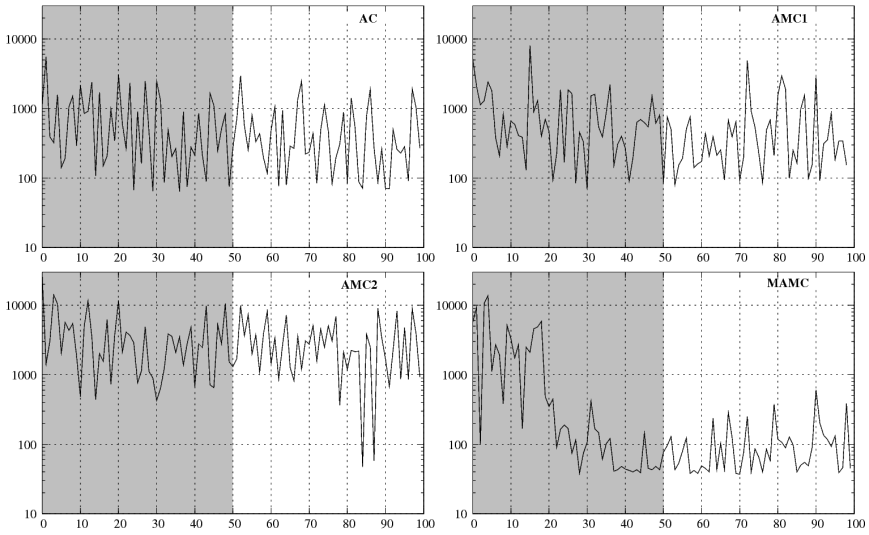

FIG. 2.5 – *Courbes d'apprentissage pour chaque modèle testé. En abcisse, les essais successifs, en ordonnée (échelle logarithmique), la durée de ces essais en pas de temps, en grisé : période d'entraînement, voir texte.*

Pour autant, ce modèle fournit des résultats bien pires (Fig. 2.4, droite, médiane significativement plus élevée, Wilcoxon : $p < 0.001$) et très instables (Fig.2.5). Il semble en effet que les discontinuités de prédiction de récompense, lors du passage du domaine d'expertise d'un critique à celui d'un autre, perturbent l'apprentissage de l'acteur unique.

Enfin, le modèle MAMC, totalement modulaire, où N couples d'acteurs-critiques se partagent le contrôle du robot à partir d'une partition a priori de l'espace d'état, est débarrassé des défauts des modèles à critiques multiples. Sa modularité augmente réellement ses capacités d'apprentissage, puisqu'il est bien plus performant que le modèle AC initial, il a même trouvé une solution au problème plus efficace que celle conçue manuellement (Fig. 2.4, droite, médiane significativement plus faible que celles de AC et GPR, Wilcoxon : $p < 0.001$).

Discussion

Ce travail montre tout d'abord qu'une tâche d'apprentissage relativement simple, dans un environnement continu, peut mettre en échec les modèles acteur-critique standards, du fait des limites computationnelles des critiques utilisés.

Cette constatation aboutit à la proposition d'un modèle modulaire, composé de plusieurs couples acteurs-critiques en charge de l'apprentissage de sous parties de la tâche globale. Cependant, ce modèle a une limite importante, puisque la répartition de ces modules dans l'espace d'état a été fournie a priori, les méthodes traditionnelles des mixtures d'expert n'ayant pas donné satisfaction. M. Khamassi a poursuivi ce travail au début de sa thèse et a proposé un modèle où cette répartition est autonome via l'utilisation de cartes auto-organisatrices (Khamassi et al., 2006), on peut donc considérer cette limite comme résolue.

Ce travail montre également qu'il est possible de réconcilier les mo-

dèles des ganglions de la base décrivant les processus de sélection et ceux intéressés à l'apprentissage. L'utilisation d'un GPR en lieu et place d'un simple WTA n'a pas empêché l'apprentissage de la tâche et a permis d'améliorer la plausibilité neurobiologique du modèle proposé. On pourra cependant regretter qu'une comparaison n'ait pas alors été réalisée entre le modèle MAMC et un modèle MAMC dont l'acteur aurait été un WTA, afin de pouvoir estimer les effets positifs ou néfastes de l'inertie de sélection d'un GPR dans le déroulement de l'apprentissage. Un point similaire a cependant été abordé depuis dans une tâche différente, relevant de la navigation, où l'acteur d'un modèle d'apprentissage par renforcement était remplacé par le modèle CBG présenté plus haut (section 2.1). Les résultats correspondant, en faveur d'un effet bénéfique de la dynamique des modèles de sélection, sont présentés en section 3.1.

2.2.2 Modulation motivationnelle

Publications : (Coninx et al., 2008)

Le travail présenté dans la section précédente correspond à un apprentissage procédural, où le système apprend à résoudre une tâche donnée, menant à un type de récompense donné. Cependant, un animal cherchant à survivre et à assurer sa descendance dans un environnement complexe, de même qu'un robot autonome idéal, assurant sa survie et l'accomplissement de diverses tâches, doit apprendre plusieurs tâches, d'une part, et apprendre à arbitrer les priorités respectives de ces tâches d'autre part. Ce second apprentissage est distinct du premier : il ne nécessite pas de réapprendre les tâches, et ne doit pas interférer avec ce premier niveau d'apprentissage.

C'est pour mettre en évidence ce point que nous avons proposé un problème-jouet de survie, similaire à la tâche précédemment utilisée, mais dans lequel la densité relative des deux sources dans l'environnement est variable[1]. D'un environnement de test à l'autre, la séquence d'actions à effectuer pour atteindre une source ne change pas : il s'agit d'explorer aléatoirement l'environnement tant que la source n'est pas visible, puis d'approcher cette source si elle entre dans le champ visuel, et de s'arrêter pour une recharge si elle est suffisamment proche. Ces actions ont des priorités croissantes, de sorte qu'au contact d'une source d'*Energie*, *ApprocheE* et *RechargeE* sont toutes deux activables, mais *RechargeE* a une saillance plus élevée et est donc sélectionnée. L'apprentissage de ces priorités relatives entre actions permettant d'obtenir un même type de récompense relève de l'apprentissage par renforcement. Cependant, dans un environnement où un type de source donné est particulièrement rare, le processus de sélection de l'action doit être biaisé de manière à favoriser l'ensemble de la séquence d'actions permettant d'atteindre cette source ; sans qu'il soit ni nécessaire, ni même utile, de remettre en cause les priorités relatives des actions dans les séquences associées à chaque source. Le contrôle de ce biais ne relève pas forcément de l'apprentissage par renforcement : la mesure des proportions relatives des sources dans l'environnement peut être effectuée de manière latente.

[1] et la consommation d'énergie des actions élémentaires augmentée à 7×10^{-3} par seconde afin rendre la tâche plus difficile à résoudre et de diminuer la durée des essais.

Modèle

Le modèle de sélection de l'action utilisé est le CBG présenté en section 2.1, modifié afin que les saillances de groupes d'actions associées à un type de source donné puisse être modulé par une estimation de la densité de ce type de sources dans l'environnement.

Les calculs de saillance du CBG original utilisés pour la résolution de la tâche de survie sont tous de la forme suivante (voir l'article, en section 6.1 pour les détails de ces calculs) :

$$\text{saillance}_i = w_i \times f(s_i) + p_i \tag{2.4}$$

où i est l'action considérée ; w_i un poids qui définit la priorité de cette action par rapport aux autres (ajusté à la main mais pouvant être l'objet d'un apprentissage par renforcement) ; s_i un produit des variables sensorielles et internes pertinentes pour l'action i ; f une fonction de transfert sigmoïde ; p_i le terme de persistance (rétroaction du CBG).

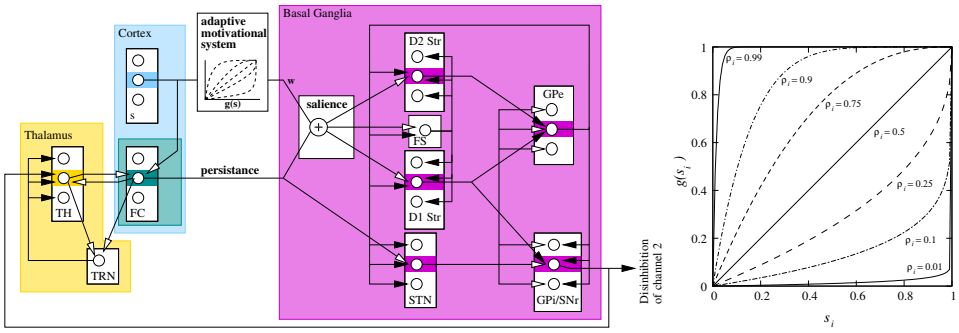

FIG. 2.6 – **Gauche :** *Modèle CBG modifié avec modulation motivationnelle adaptative. Les saillances passent dans une fonction de transfert g dont la non-linéarité paramétrable est adaptée en fonction des conditions environnementales perçues par le système.* **Droite :** *Fonction de transfert $g(s_i, \rho_i)$, représentée pour différentes valeurs de ρ_i.*

La modulation proposée consiste à remplacer f par une fonction de transfert g, initialement proposée dans Konidaris et Barto (2006), de la forme (Fig. 2.6, gauche) :

$$g(s_i, \rho_i) = 1 - (1 - s_i)^{tan\frac{\rho_i \pi}{2}} \tag{2.5}$$

sa non-linéarité dépend du paramètre ρ_i (voir Fig. 2.6, droite).

Les ρ_i sont initialisés à des valeurs permettant d'assurer des performances similaires à celles obtenues avec le CBG original dans l'environnement de référence (une source de chaque type dans un environnement de 10×10 mètres, voir annexe A pour le détail de ces valeurs). Une variation de la densité estimée d'une source dans l'environnement va entraîner une variation de signe opposé des ρ_i des actions i dirigées vers cette source (actions dites « appétitives »). Ainsi, si une source est plus abondante que la référence, toutes les actions menant à cette source vont voir leurs saillances modulées à la baisse, sans que leurs priorités relatives ne soient affectées. Enfin, l'action de repos voit son ρ_i augmenter avec l'abondance de l'une ou l'autre des sources, puisque cette action ne doit être favorisée que dans les environnements riches en sources.

L'estimation de la densité des sources dans l'environnement est simplement fondée sur le calcul d'une moyenne glissante du nombre d'apparitions dans le champ visuel sur une fenêtre temporelle suffisament longue (25 secondes dans notre cas, voir annexe A pour l'algorithme). Ces mises à jour sont effectuées en continu et peuvent donc s'adapter à des condition environnementales changeant au cours de la vie de l'animat.

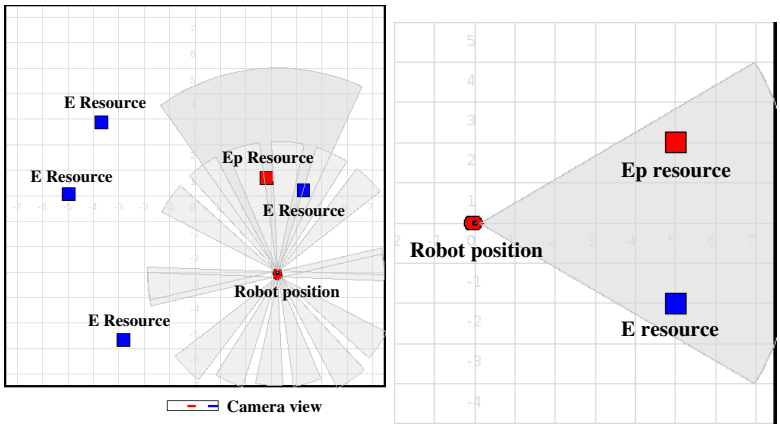

FIG. 2.7 – *Environnements de test.* **Gauche :** *environnement d'entraînement, configuration* $(4E, 1E_p)$ *;* **Droite :** *test du choix comportemental après entraînement.*

Résultats

L'efficacité du mécanisme proposé a été évaluée dans la tâche de survie, dans un environnement de 15×15 mètres contenant de 1 à 4 exemplaires de chaque source, placés aléatoirement (distribution uniforme à plus d'un mètre des murs, exemple Fig. 2.7, gauche). Les 5 conditions testées sont : (1 E, 1 E_p), (1 E, 2 E_p), (1 E, 4 E_p), (2 E, 1 E_p), (4 E, 1 E_p). Pour chacune de ces conditions, 50 dispositions de sources ont été tirées. Pour chacune de ces 50 dispositions, un robot simulé contrôlé par un CBG, avec et sans mécanisme d'adaptation motivationnelle, a effectué la tâche de survie. Le robot ayant initialement $E = 1$ et $E_p = 0$, sa durée de vie minimale est de $2min$ $23s$, la durée maximale d'un essai est fixée à $30min$.

TAB. 2.1 – *Nombre d'essais réussis et médianes des consommation d'énergie pour les systèmes avec (A) et sans (NA) motivation adaptative, pour chaque condition.*

		$(1E, 1E_p)$	$(1E, 2E_p)$	$(1E, 4E_p)$	$(2E, 1E_p)$	$(4E, 1E_p)$
Succès	A	18	25	29	35	33
	NA	11	27	34	38	37
Conso	A	6.2	6.0	5.6	**5.4**	**4.9**
$(10^{-3}.s^{-1})$	NA	6.3	5.9	5.7	**6.1**	**6.0**

Il n'y a pas de différences notables des durées de survies entre les systèmes avec et sans adaptation. Le nombre de succès ($30min$ de survie) est relativement similaire (Tab. 2.1), on peut noter un avantage assez net du système adaptatif dans la situation de référence (18 contre 11), et un léger avantage du système non adaptatif dans les autres situations. Les

distributions de durée de vie des individus n'ayant pas survécu 30*min*, ne sont cependant pas significativement différentes (test KS).

En revanche, grâce à une utilisation plus importante de l'action de repos, la consommation moyenne d'énergie est significativement plus faible dans les environnements riches en E. Le tableau 2.1 indique la médiane de la consommation moyenne par essai, pour les 50 essais de chaque combinaison système×environnement. Dans les cas où il y a 2 ou 4 sources d'*nergie*, les médianes avec et sans adaptation sont significativement différentes (test de Wilcoxon, indiquées en gras). Cet effet ne peut pas être constaté dans les environnements riches en E_p du fait du rôle non symétrique des deux sources dans le métabolisme virtuel (voir Coninx et al., 2008, pour une explication détaillée).

Un test de choix comportemental est de plus effectué après entraînement, en utilisant la moyenne des ρ_i obtenus dans les 50 essais d'une même condition environnementale. Il s'agit d'évaluer l'incidence de l'adaptation réalisée dans chaque condition sur un choix de type « âne de Buridan » (représenté sur la Fig. 2.7, droite). Le robot est positionné à égale distance d'une source visible de chaque type, on enregistre la première source où il se recharge dans les 30 premières secondes de la simulation, pour toutes les combinaisons de valeurs de E et E_p entre 0.1 et 1 avec un pas de 0.1. On constate (Fig. 2.8) qu'en fonction de l'environnement auquel a été exposé l'animat, la frontière séparant les choix entre les deux types de source se déplace en faveur de la source la moins abondante.

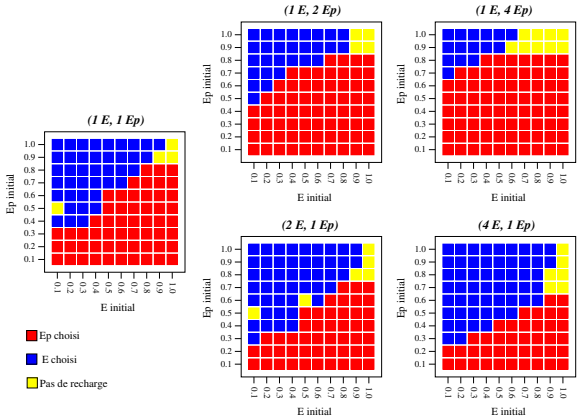

FIG. 2.8 – *Source choisie dans le test de choix comportemental, en fonction des valeurs initiales de E et de E_p, après adaptation à diverses densités de sources. Le déplacement de la frontière de choix entre bleu et rouge se fait en faveur de la source la moins abondante.*

Discussion

Cette étude n'est pas totalement mature et n'a d'ailleurs pas donné lieu à une publication de journal. Elle est cependant incluse dans ce manuscrit afin d'illustrer l'idée de fond que la sélection de l'action nécessite, certes, d'être paramétrée par apprentissage par renforcement, mais que dans le cadre de processus motivationnels concurrents, elle peut aussi nécessiter d'être modulée via un apprentissage latent.

Les résultats obtenus ici sont empreints d'une grande variabilité, et les avantages éventuels, en terme de survie, du mécanisme de modulation

motivationnel proposé ne sont pas clairement avérés. Cela est certainement du à l'algorithme *ad hoc* utilisé, qui mériterait d'être amélioré et testé plus avant. Par exemple, le calcul de ρ pour l'action de repos devrait plutôt résulter d'un produit que d'une somme des observations de chaque type de source, car dans un environnement très riche d'un seul type de source, cela risque d'entraîner trop d'arrêts, au détriment du type de source rare.

D'un point de vue neurobiologique, cette modulation pourrait résulter de l'action des circuits ventraux des ganglions de la base, en particulier celui issu de la partie externe (*shell*) du noyau Accumbens (Kelley, 1998, 1999), sur la base de son rôle dans la théorie de l'*incentive salience* de Berridge et Robinson (1998). En effet, ce circuit est en situation de moduler l'activité des circuits dorsaux via ses projections dopaminergiques (Joel et Weiner, 2000). Sur la base de cette hypothèse de substrat neural, ainsi que sur celle des similarités entre circuits ventraux et dorsaux des ganglions de la base, il semble possible de poursuivre ces travaux par le développement d'un modèle neuromimétique complet de modulation par l'état motivationnel des processus de sélection.

2.3 Modèle bayésien de sélection

Publications : (Colas et al., 2008, 2009)

L'accumulation de données expérimentales montrant la prise en compte de l'incertitude des informations sensorielles par les processus délibératifs du système nerveux central, a favorisé l'émergence de nombreux modèles computationnels fondés sur des approches bayésiennes. Parmi celles-ci, la programmation bayésienne (Bessière et al., 2008) a l'avantage de proposer un cadre formel englobant de nombreuses méthodes probabilistes classiques (réseaux bayésiens, modèles de markov cachés, filtres de Kalman, etc., voir Bessière et al., 2003). La programmation bayésienne a ainsi permis, dans le cadre des neurosciences, la conception de modèles de l'intégration des informations vestibulaires(Laurens et Droulez, 2007) ou encore de la perception 3D à partir du flux optique (Colas et al., 2007). C'est dans le cadre d'une collaboration avec P. Bessière (LIG) et deux postdocs (F. Flacher et F. Colas), pour la partie modélisation, ainsi qu'avec T. Tanner (MPI, Tübingen), pour la partie expérimentale, que nous avons étudié l'importance de la prise en compte de l'incertitude dans la sélection de l'action. Pour ce faire, nous avons cherché à modéliser le comportement de sujets humains dans une tâche de psychologie expérimentale nécessitant des procéder à des sélections.

La tâche proposée aux sujets est une modification de la tâche de suivi d'objets multiples (MOT, Pylyshyn et Storm, 1988). Dans cette tâche, le sujet doit ancrer son regard sur un point de fixation au centre d'un écran, il se voit présenter un ensemble de stimuli correspondant à des cibles et des distracteurs, les cibles étant identifiées en début de tâche par un clignotement (voir Fig. 2.9). Ces objets se mettent ensuite en mouvement, alors que cibles et distracteurs redeviennent identiques. Le sujet doit donc retenir la position initial des cibles et suivre leur mouvement. En fin de tâche, le mouvement s'arrête et le sujet doit désigner parmi les objets ceux qu'il pense être les cibles initialement présentées. Notre version de la tâche

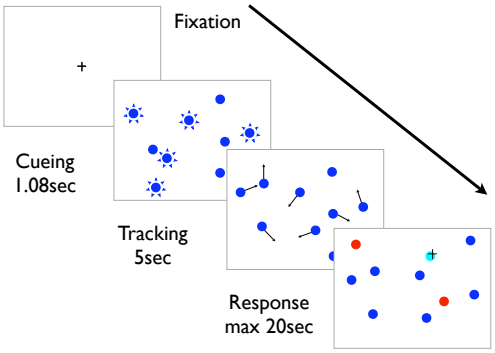

FIG. 2.9 – *Tâche de suivi de cibles multiples. Les cibles clignotent durant* 1.08*s, le mouvement dure ensuite* 5*s, puis le sujet doit désigner la nouvelle position des cibles en moins de* 20*s.*

a consisté à autoriser les mouvement des yeux dans la moitié des essais, plutôt que de forcer un ancrage sur le point de fixation. Le sujet est donc susceptible d'orienter sa fovéa vers les régions de l'écran susceptible de lui fournir un maximum d'information pour la résolution du problème, par exemple vers les configurations où une cible passe à proximité d'un distracteur. Ce sont les critère de choix utilisés pour orienter le regard à chaque instant qui nous ont intéressé.

La méthodologie de la programmation bayésienne nous a quelque peu éloigné des modèles neuromimétiques utilisés dans l'ensemble des autres travaux présentés dans ce manuscrit, pour autant, la structure du modèle proposé emprunte nombre de caractéristiques aux régions du cerveau impliquées dans les mouvements des yeux. Ainsi, il est composé de cartes rétinotopiques à la géométrie complexe-logarithmique spécifique du colliculus supérieur et du cortex visuel. Par ailleurs, ces cartes ont des activités visuelles, mémorielles ou motrices, reproduisant en cela les différentes catégories de neurones identifiées dans le SC (voir section 1.3.2), mais également dans les champs oculaires frontaux et le cortex pariétal.

2.3.1 Modèle

Le modèle proposé est constitué de deux niveaux : un premier de *représentation* des informations en provenance du champ visuel, et un second de *décision* du prochain mouvement oculaire.

Représentation

La partie *représentation* du modèle est composée de cartes rétinotopiques dynamiques (au sens de Droulez et Berthoz, 1991), utilisant la géométrie complexe-logarithmique des cartes colliculaires. Ces cartes sont regroupées en deux couches principales : l'*occupation* du champ visuel (par des objets, cibles ou distracteurs) et des mémoires de la *position* de chaque cible.

La carte d'occupation est une simple grille d'occupation, un filtre Bayésien récursif initialement utilisé pour la représentation de la position des obstacles en robotique mobile (Elfes, 1989). L'environnement (ici le champ

visuel) est discrétisé en une grille régulière $\mathcal{G}$, et nous définissons une variable binaire Occ_c^t pour chaque élément c de la grille, à chaque pas de temps t, qui indique si un objet est présent dans la région correspondante du champ visuel. Les observations visuelles sont également un ensemble de variables binaires Obs_c^t, et $P(Obs_c^t \mid Occ_c^t)$ représente la probabilité d'observer l'occupation d'une cellule supposée occupée. Cette distribution de probabilité, ainsi que les suivantes dans le modèle, a une forme paramétrique donnée, dont les paramètres sont appris à partir des données expérimentales.

La capacité de mise à jour de l'occupation du champ visuel après un mouvement oculaire Mvt^t est fondée sur la distribution $P(Occ_c^t \mid Occ_{\mathcal{A}(c)}^{t-1}\ Mvt^t)$ qui transfère l'occupation de l'ensemble $\mathcal{A}(c)$ des cellules antécédentes à la cellule courante par le mouvement Mvt, avec une incertitude additionnelle. La mise à jour de la connaissance du modèle de l'occupation du champ visuel est opérée récursivement :

$$\begin{aligned} & P(Occ_c^t \mid Obs^{1:t}\ Mvt^{1:t}) \\ \propto\ & P(Obs_c^t \mid Occ_c^t) \\ & \times \sum_{Occ_{\mathcal{A}(c)}^{t-1}} \left[\begin{array}{l} P(Occ_c^t \mid Mvt^t\ Occ_{\mathcal{A}(c)}^{t-1}) \\ \prod_{c'} P(Occ_{c'}^{t-1} \mid Obs^{1:t-1}\ Mvt^{1:t-1}) \end{array} \right. \end{aligned} \tag{2.6}$$

Pour discriminer les cibles des distracteurs, le modèle dispose d'un ensemble de variables Tgt_i^t qui représentent la position de la cible i à t. Ces représentation sont dotées des mêmes capacités de mise à jour que la grille d'occupation, grâce à un modèle dynamique $P(Tgt_i^t \mid Tgt_i^{t-1}\ Occ^t\ Mvt^t)$ similaire. La mise à jour de la connaissance du modèle sur les cibles est donc calculée à chaque pas de temps ainsi :

$$\begin{aligned} & P(Tgt_i^t \mid Obs^{1:t}\ Mvt^{1:t}) \\ \propto\ & \sum_{Tgt_i^{t-1}} \left[\begin{array}{l} P(Tgt_i^{t-1} \mid Obs^{1:t-1}\ Mvt^{1:t-1}) \\ \times \sum_{Occ^t} \left[\begin{array}{l} P(Occ^t \mid Obs^{1:t}\ Mvt^{1:t}) \\ \times P(Tgt_i^t \mid Tgt_i^{t-1}\ Occ^t\ Mvt^t) \end{array} \right. \end{array} \right. \end{aligned} \tag{2.7}$$

Les questions 2.6 et 2.7 sont la connaissance courante qu'a le modèle de la scène visuelle, par inférence sur les mouvements et observations passés.

Décision

A partir de cette connaissance, nous proposons trois modèles devant décider du prochain mouvement. Pour vérifier que le modèle de représentation que nous avons choisi est utile, nous comparons un modèle sans cette représentation (modèle *constant*) avec un modèle minimal la possédant (modèle *cibles*). Enfin, l'hypothèse principale de ce travail étant que l'incertitude sur la position des cibles est prise en compte dans la décision de mouvement, nous comparons le modèle *cibles* avec un modèle prenant explicitement en compte l'incertitude (modèle *incertitude*).

Le modèle *constant* est défini comme la meilleure distribution de probabilité statique $P(Mot)$ pouvant rendre compte des mouvements mesurés expérimentalement. Dans cette distribution, la probabilité d'un mouvement donné est égal à sa fréquence expérimentale.

Le modèle *cibles* est un modèle de fusion Bayésienne où la position de chaque cible en mémoire est considérée comme la cible possible du

prochain mouvement. C'est un modèle inverse $P(Tgt_i^t \mid Mot^t)$ qui postule qu'à t, la position de la cible Tgt_i^t est probablement proche du prochain mouvement de l'œil Mvt^t, avec une distribution Gaussienne. De plus, la distribution a priori est celle du modèle constant. Par conséquent, ce modèle affine la distribution de mouvement des yeux en prenant en compte l'influence des positions courantes des cibles. Comme cette position exacte n'est pas connue, ce modèle utilise l'estimation de la question 2.7 pour réaliser la fusion :

$$\begin{aligned} & P(Mot^t \mid Obs^{1:t}\ Mvt^{1:t}) \\ & \propto\ P(Mot) \prod_{i=1}^{N} \sum_{Tgt_i^t} P(Tgt_i^t \mid Obs^{1:t}\ Mvt^{1:t}) P(Tgt_i^t \mid Mot^t) \end{aligned}$$

Sans manipuler explicitement l'incertitude, le modèle *cibles* est influencé par elle en ce que l'incitation à regarder une cible donnée est plus forte si la position de cette cible est connue avec plus de certitude. Dans le modèle *incertitude*, nous proposons d'inclure l'incertitude en tant que variable sur laquelle raisonner : comme la connaissance à décrire. L'idée étant qu'il semble plus efficace d'acquérir de l'information quand et là où elle manque, c'est-à-dire quand et où il y a le plus d'incertitude. Nous introduisons donc un nouvel ensemble de variables I_c^t représentant l'index d'incertitude de la cellule c à l'instant t. Nous le spécifions en fonction de la distribution de probabilité de l'occupation dans cette cellule. Plus cette probabilité est proche de 0.5, plus grande est l'incertitude et la probabilité de regarder là. La distribution de probabilité a posteriori du prochain mouvement est calculée comme suit :

$$\begin{aligned} & P(Mot^t \mid Obs^{1:t}\ Mvt^{1:t}\ I^{1:t}) \\ & \quad \propto\ \ P(Mot^t \mid Obs^{1:t}\ Mvt^{1:t}) P(I^t_{Mot^t} \mid Mot^t) \end{aligned}$$

avec $I_c^t = P(Occ_c^t \mid Obs^{1:t}\ Mvt^{1:t})$ (équation 2.6). Le modèle *incertitude* filtre donc la distribution de mouvements calculée par le modèle *cibles*, afin de l'amplifier dans les régions de forte incertitude.

2.3.2 Résultats

Les données expérimentales on été collectées sur 11 sujets, chacun exécutant 110 essais, pour un total de 1210 essais (voir Tanner et al., 2007, pour plus de détails). Chaque essai comprend 24 observations, ce qui donne un total de 29040 mesures. 124 essais choisis au hasard on été utilisés pour déterminer les 9 paramètres du modèle, et les résultats on été calculés sur les 1089 essais restants.

Les trois modèles de décision produisent des distributions de probabilité du prochain mouvement à chaque pas de temps (Fig. 2.10), qui dépendent des mouvements et observations passés, dans le référentiel rétinocentrique (à l'exception du modèle *constant*). Nous pouvons donc calculer et comparer à chaque pas de temps la probabilité, pour chaque modèle, du mouvement réellement effectué (estimation du maximum de vraisemblance). Afin d'effectuer cette comparaison sur l'ensemble des données mesurées avec une mesure ne tendant pas vers 0 avec l'augmentation du nombre de mesures, nous comparons les moyennes géométriques des vraisemblances par essai (Tab. 2.2).

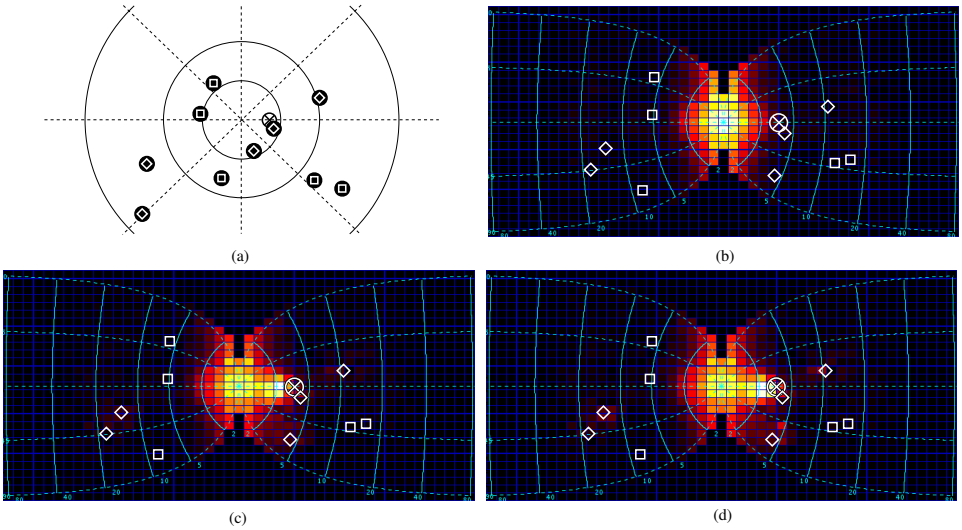

FIG. 2.10 – *Exemples de distributions de probabilités calculées par chaque modèle dans la même configuration. (a) configuration des cibles ◇ et distracteurs □ dans le champ visuel (⊗ prochain mouvement mesuré expérimentalement) ; (b) modèle* constant *; (c) modèle* cibles *; (d) modèle* incertitude.

Ratio	Constant	Cibles	Incertitude
Constant	1	280	320
Cibles	3.5×10^{-3}	1	1.14
Incertitude	3.1×10^{-3}	0.87	1

TAB. 2.2 – *Ratio des moyennes géométriques des vraisemblances par paires de modèles.*

2.3.3 Discussion

Nous avons donc proposé un modèle bayésien intégrant explicitement l'incertitude dans son processus computationnel (plutôt que l'utilisant implicitement, comme c'est en général le cas), et nous avons montré qu'il expliquait mieux les données expérimentales qu'un modèle purement statistique, ou qu'un modèle ne prenant en compte que la position estimée des cibles. Bien que ne proposant pas un modèle détaillé de la façon dont les circuits neuronaux du contrôle des mouvements des yeux prennent en compte l'incertitude, cette étude met donc en avant le fait qu'ils le font très probablement. Il est par ailleurs intéressant de noter que les modèles *constant* et *incertitude* ont une vraisemblance similaire si l'on ne considère que les petits mouvements relevant de la fixation ou de la poursuite lente, et que la différence entre ces deux modèle provient donc essentiellement des mouvements saccadiques.

La transcription sous forme de réseaux de neurones de ce modèle, par exemple sur la base des modèles proposés par Rao (2004) ou plus récemment Denève (2008a,b), permettrait peut-être de faire le lien avec les travaux de modélisation des ganglions de la base et du colliculus que j'ai mené par ailleurs (voir respectivement sections 2.1 et 4.1), qui sont bien plus proches des données neurobiologiques, mais qui négligent, en l'état, la présence d'incertitude dans les données sensorielles.

2.4 Discussion générale

Les trois premiers travaux présentés participent à un objectif commun : la modélisation des boucles cortico-baso-thalamo-corticales en intégrant leurs rôles dans la sélection de l'action et dans l'adaptation de la sélection. Dans chaque cas, les modèles ont été évalués dans des tâches relativement simples, dont le but n'est naturellement pas de modéliser avec précision les problèmes de survie auquel peut être confronté un animal, mais de mettre en exergue des propriétés des modèles proposés : l'apprentissage à partir d'entrées continues et avec un grain temporel fin, la limitation ses oscillations comportementales, etc. Cette approche semble inévitable dans la mesure où, d'une part, il est difficile d'établir des protocoles expérimentaux permettent un suivi en continu du comportement chez les animaux sur des périodes de temps suffisamment longues pour établir une comparaison directe avec les modèles de sélection de l'action et où, d'autre part, des modèles quantitatifs du métabolisme de ces animaux seraient nécessaires.

A ce titre, il est intéressant de constater que l'ajout de capacités d'apprentissage par renforcement aux modèles des BG, même dans les tâches relativement simplistes où nous les testons, pose des problèmes. Ainsi, la tâche de survie minimale que nous utilisons nécessite le calcul de saillances combinant les entrées sensorielles par des sommes pondérées, mais également par des produits. La découverte automatique de telles combinaisons de variables pertinentes n'est pas triviale et sort du domaine des algorithmes d'apprentissage, relativement simples, qui ont jusqu'ici été importés de l'apprentissage automatique vers les neurosciences.

Enfin, une question se pose en filigrane de l'ensemble des travaux exposés dans ce chapitre, y compris le dernier ; elle concerne l'exploration, ou, plus précisément, l'implémentation neuronale des processus de tirage aléatoire dans des distributions. En effet, les algorithmes d'apprentissage par renforcement, tel l'acteur-critique utilisé pour modéliser les ganglions de la base, nécessitent pour converger qu'un processus d'exploration pousse régulièrement le système à ne pas effectuer le meilleur choix au vu de ses connaissances actuelles. Une solution classique à ce problème consiste à choisir l'action sur la base d'un tirage dans une distribution de probabilités $P(a_i)$, résultant par exemple d'un *softmax* appliqué aux sorties $O(a_i)$ de l'acteur : $P(a_i) = e^{\beta O(a_i)} / \sum_j e^{\beta O(a_j)}$. Le contraste y est contrôlé par le paramètre β : pour β faible, la distribution résultante gomme les différences entre les valeurs des actions, poussant à l'exploration, alors qu'avec un β fort, seule l'action dont la sortie est la plus importante à de réelles chances d'être sélectionnée. Si on explore la possibilité qu'un algorithme similaire soit implémenté dans le cerveau, les ganglions de la base devraient-ils être en charge de réaliser l'opération de modulation du contraste, auquel cas leur sortie devrait être interprétée comme une distribution de probabilités et le tirage dans cette distribution serait effectué dans un de leurs circuits cibles ? Ou bien réaliseraient-ils eux-même le tirage sur la base d'une distribution calculée en entrée, par exemple dans le striatum ? Dans ce dernier cas, pour permettre le tirage d'actions diverses pour une même distribution, l'instabilité du système serait souhaitable et la propriété de contraction probablement à éviter.

Navigation

3

Les animaux ont à leur disposition de nombreuses stratégies pour résoudre les problèmes de navigation auxquels ils sont continuellement confrontés. Ils sont en particulier capables de changer de stratégie chaque fois que les circonstances l'exigent. Ces stratégies peuvent être très variées (voir Redish, 1999; Arleo et Rondi-Reig, 2007; Khamassi, 2007, pour des revues détaillées de ces stratégies), allant des plus simples, ne nécessitant pas de construction d'une représentation de l'environnement (exploration, approche d'objet, intégration de chemin), aux plus complexes (construction d'une représentation topologique ou métrique de l'espace et planification de trajectoire dans ces représentations). Il semble que les substrats neuraux de ces stratégies, au moins pour celles fondées sur de simples apprentissages stimulus-réponse (S-R) et pour celles requérant l'usage d'une représentation topologique de l'espace (carte cognitive), soient distincts (Packard et al., 1989; McDonald et White, 1993; Devan et White, 1999; Kim et Baxter, 2001; White et McDonald, 2002; Burgess, 2008).

Ces stratégies de navigation nécessitent des algorithmes fondamentalement différents et semblent mettre à contribution des substrats neuronaux distincts. Enfin, la dynamique de leurs interactions (compétition, coopération) n'est pas encore bien comprise et reste un sujet de recherche extrêmement actif.

Dans un premier travail mené à la fin de ma thèse (Girard et al., 2004, 2005a), j'ai proposé une architecture neuromimétique fondée sur deux circuits parallèles des ganglions de la base, réalisant la fusion de deux stratégies de navigation (approche visuelle d'objets et planification de trajectoire dans une carte topologique établie par exploration autonome). L'efficacité de cette architecture a pu être établie en la soumettant à une contrainte de survie similaire à celle utilisée en sélection de l'action (voir section 2.1). Cependant, l'importance relative de chaque stratégie dans le processus de fusion était fixé *a priori* : il était possible de concevoir un agent préférant l'une ou l'autre des stratégies en situation de choix, mais il ne pouvait ensuite modifier cette préférence de manière autonome, en fonction de l'environnement auquel il était confronté.

J'ai donc contribué à l'élaboration par L. Dollé et D. Sheynikhovich, supervisés de A. Guillot et R. Chavarriaga, d'un modèle d'apprentissage de la sélection de stratégies de navigation, capable de reproduire des données comportementales chez le rat en situation de conflit (3.1). Afin d'enrichir le répertoire de stratégies de navigation modélisées et donc d'élargir la liste des expériences reproductibles, j'ai également encadré les travaux de C. Masson portant sur le processus d'intégration de chemin permettant le

retour au point de départ (*homing*). Plus précisément, il s'agissait d'implémenter un réseau de neurones capable d'extraire de l'activité des cellules de grilles du cortex entorhinal, les coordonnées de l'animal (3.2).

3.1 Apprentissage et Sélection de stratégies multiples

Publications : (Dollé et al., 2008, 2010a,b)

Les stratégies S-R sont contrôlées par les circuits des BG issus du striatum dorso-latéral (DLS) et semblent résulter d'un apprentissage par renforcement relativement lent et inflexible. Elles associent en général des indices sensoriels proches (par exemple une cible visuelle indiquant la position d'une plate-forme immergée) avec des actions (s'orienter vers cette cible). Les stratégies utilisant une carte cognitive impliquent l'hippocampe, le cortex préfrontal et probablement les circuits des BG issus du striatum dorso-médian (DMS) ; elles sont apprises rapidement et sont flexibles, par exemple lors de changements environnementaux. La carte cognitive semble essentiellement fondée sur l'activité des cellules de lieux de l'hippocampe, qui s'activent lorsque l'animal est dans une zone restreinte de l'environnement . Cette activité particulière résulte principalement de la fusion de l'apprentissage des configurations d'indices visuels lointains en un lieu donné avec l'intégration des mouvements propres permettant de positionner ce lieu relativement aux autres.

Les interactions de ces deux types de stratégies ont souvent été analysées en terme de coopération (lorsque la lésion du substrat neural de l'un des stratégies dégrade les performances de celle qui reste) ou de compétition (lorsque la lésion du substrat de l'une des stratégies améliore l'apprentissage de l'autre). Ces deux types d'effets sont observables en fonction du protocole expérimental mis en œuvre, de sorte que le mécanisme d'interaction entre stratégies est encore actuellement mal compris.

C'est pourquoi nous avons cherché à répondre, par le biais d'un modèle computationnel de l'apprentissage simultané de deux stratégies et de leur sélection, aux questions suivantes :

1. Quel mécanisme de sélection de stratégies peut expliquer la compétition ou la coopération entre stratégies, observées expérimentalement ?
2. Quel critère est utilisé par ce mécanisme pour choisir entre des stratégies dont les algorithmes sont, *a priori*, totalement différents ?
3. Comment les informations sensorielles proches et lointaines sont-elles susceptibles d'affecter cette sélection ?

3.1.1 Modèle

Le modèle proposé est un modèle de *sélection* de stratégies multiples (voir Fif. 3.1, gauche), appliqué pour l'instant à une stratégie d'approche d'indice visuel (*taxon*), à une stratégie de *planification* de chemin dans une carte cognitive et à une stratégie de déplacement aléatoire (*exploration*). Il est fondé sur trois hypothèses principales :

1. Les stratégies de navigation considérées ont des substrats neuraux différents qui apprennent simultanément et indépendamment les

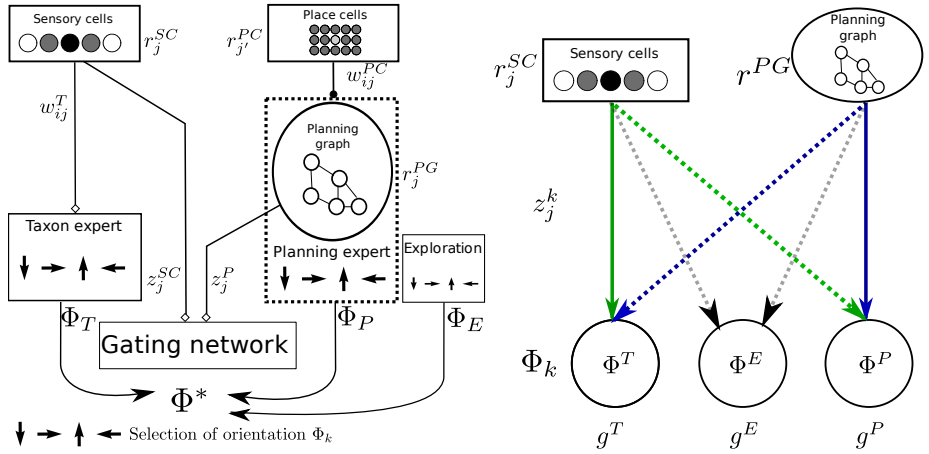

FIG. 3.1 – ***Gauche :*** *Structure du modèle : le* taxon *prend ses décisions sur la base des informations visuelles signalant la présence d'indices proches ; la* planification *construit un graphe sur la base de l'activité des cellules de lieux et des déplacements effectués, puis l'utilise pour planifier des trajectoires ; l'*exploration *génère des suggestions de déplacements aléatoires ; le réseau de sélection choisit la stratégie qui contrôle les déplacements Φ^* à chaque pas de temps, en fonction des données sensorielles et des nœuds actifs du graphe.* ***Droite :*** *Detail du circuit de sélection : l'attribution à chaque stratégie d'une valeur de choix g^{strat} résulte de l'apprentissage par renforcement des poids z d'un réseau de neurones à une couche, prenant en entrée les mêmes données que celles utilisées par les stratégies pour prendre leurs décisions.*

uns des autres, comme dans les modèles de Guazzelli et al. (1998); Girard et al. (2005a); Chavarriaga et al. (2005).

2. Les processus d'apprentissage et les algorithmes de ces stratégies sont différents : le *taxon* utilise de l'apprentissage par renforcement pour apprendre des associations stimulus-réponse immédiatement utilisables, là où la *planification* apprend de manière latente les liens entre lieux[1] et utilise une recherche dans un graphe (par propagation d'activité) pour générer ses réponses. Ces différences algorithmiques sont aussi présentes dans les modèles de Guazzelli et al. (1998); Girard et al. (2005a).
3. Le mécanisme de sélection de stratégie est adaptatif : comme dans le modèle de Chavarriaga et al. (2005), il est capable de se mettre automatiquement à jour en fonction des modifications environnementales. Il apprend par renforcement à associer à chaque contexte (entrées sensorielles et activité des cellules de lieux) la stratégie la plus efficace en terme de récompense obtenue (voir Fig. 3.1, droite). Ses originalités sont que, d'une part, bien qu'une seule stratégie soit sélectionnée à un moment donné pour contrôler la direction du déplacement, toutes les stratégies peuvent utiliser la différence entre la direction qu'elles proposaient et celle effectivement choisie pour apprendre. D'autre part, il utilise les mêmes informations sensorielles que les stratégies en plus des directions de déplacement qu'elles suggèrent. Il est donc indépendant des algorithmes mis en œuvre par

[1] Dans(Dollé et al., 2010a) des cellules de lieux *ad hoc* sont fournies au modèle, alors que dans (Dollé et al., 2010b), elles sont générées par le modèle d'hippocampe de (Ujfalussy et al., 2008)

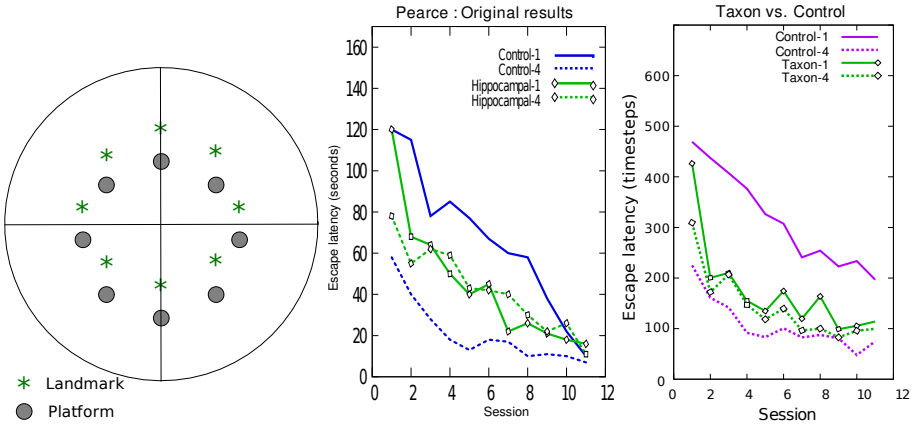

FIG. 3.2 – ***Gauche :*** *Dispositif expérimental de Pearce et al. (1998). Les étoiles représentent la position de l'indice visuel par rapport à la plate-forme (disque gris).* ***Milieu :*** *Résultats originaux.* ***Droite :*** *Résultats de la simulation.*

ces stratégies, là où le modèle de Chavarriaga et al. (2005) ne peut que coordonner des stratégies utilisant de l'apprentissage par renforcement.

Les algorithmes utilisés sont détaillés dans l'article (Dollé et al., 2010a), inclus en section 6.4.

3.1.2 Résultats

Deux expériences comportementales avec lésions ont été simulées avec ce modèle : celle de Pearce et al. (1998) et celle de Devan et White (1999).

Pearce et al. (1998)

Dans cette expérience, un groupe de rats contrôle et un groupe ayant une lésion du fornix (voie de sortie de l'hippocampe) doivent apprendre à retrouver la position d'une plate-forme immergée dans une piscine circulaire (voir Fig. 3.2, gauche). Pour ce faire, ils peuvent utiliser des indices distants pour construire une représentation de l'environnement, s'y localiser et apprendre où se trouve la plate-forme, ou bien utiliser un indice visuel proche, toujours placé à la même distance de la plate-forme, dans une même direction absolue. Le protocole expérimental définit des 12 sessions composées chacune de 4 essais. La plate-forme et l'indice sont déplacés au début de chaque session.

Les principaux résultats expérimentaux obtenus (voir Fig. 3.2, milieu) sont les suivants : les deux groupes de rats apprennent à retrouver le but ; les rats avec lésion du fornix trouvent plus rapidement le but que les rats contrôles au premier essai d'une session, alors que cette tendance est inversée au quatrième essai ; les performances des deux groupes s'améliorent de session en session. La simulation du modèle de sélection de stratégie proposé génère des résultats similaires (voir Fig. 3.2, droite), si l'on simule la lésion du fornix par la suppression du module de *planification*. La seule différence notable est que les performances du groupe

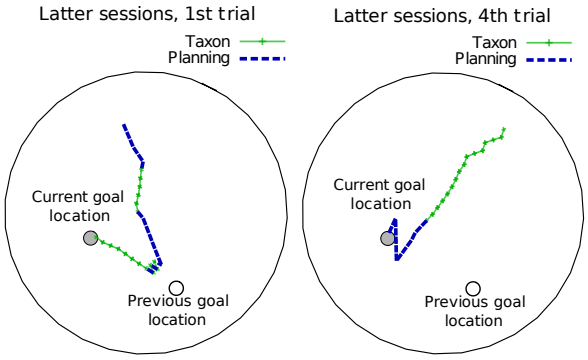

FIG. 3.3 – ***Gauche :*** *Trajectoire d'un rat simulé au premier essai d'une session de fin d'expérience. La* planification *oriente vers l'ancienne position de la plate-forme et perturbe l'exécution de la stratégie* taxon. ***Droite :*** *Trajectoire pour le quatrième essai d'une session de fin d'expérience. Le* taxon *contrôle la direction générale du déplacement en début de trajectoire, alors que la* planification *se charge de la phase finale d'approche.*

contrôle au premier essai de la douzième session ne sont pas encore identiques à celles du quatrième essai.

L'analyse de ces simulations permet de constater que la mauvaise performance du groupe contrôle au premier essai des sessions est due au fait que la stratégie de *planification* ne sait pas encore que la plate-forme a été déplacé, et pousse donc à approcher l'ancienne position (Fig. 3.3, gauche). Il s'agit là d'un effet de compétition entre stratégies observé entre deux sessions, compatible avec les interprétations proposées par Pearce et al. (1998). Cependant, durant une session, lorsque la plate-forme ne change pas de position, les deux stratégies coopèrent : les performances du groupe contrôle deviennent meilleures que celles du groupe lésé. Dans le cadre des simulations, cela résulte de l'utilisation des deux stratégies dans les régions où elles sont les plus efficaces : le *taxon* est utilisé en début de trajectoire pour se diriger dans la bonne région de la piscine, alors que la *planification* se charge de la fin de la trajectoire, là où le *taxon* a des difficultés à guider l'animal simulé vers la bonne position relative à l'indice visuel (Fig. 3.3, droite).

Devan & White (1999)

Dans cette expérience, quatre groupes de rats (contrôles, lésion du fornix, du DLS, du DMS) doivent apprendre à trouver une plate-forme au cours de neuf sessions, sachant qu'elle est dissimulée sous l'eau aux sessions 3, 6 et 9. Un test de compétition entre stratégies est effectué en fin d'expérience (session 10), la plate-forme est cette fois visible mais située à un autre emplacement.

Les résultats obtenus pour les lésions du DMS sont difficilement transcriptibles dans le cadre de notre modèle : résultent-ils d'une perturbation de la stratégie de *planification* ? D'une perturbation du mécanisme de sélection ? Dans ce dernier cas, à quel niveau ? Ces incertitudes nous ont mené à nous limiter dans cette étude aux trois autres groupes, pour lesquels les résultats expérimentaux (Fig. 3.4, gauche) sont les suivants : les trois groupes apprennent à rejoindre la plate-forme visible avec des per-

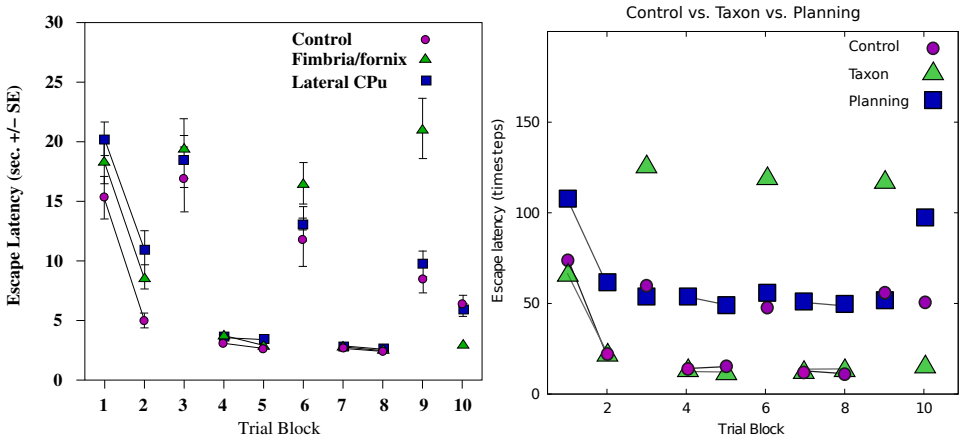

FIG. 3.4 – ***Gauche :*** *Résultats originaux de (Devan et White, 1999) sur les trois groupes contrôle, lésion du fornix et lésion du DLS.* ***Droite :*** *Résultats de la simulation.*

formances similaires, car les deux stratégies permettent de trouver le but dans cette configuration ; les rats avec une lésion du fornix sont plus lents à trouver la plate-forme dans les essais où elle est cachée, car seule l'utilisation de la carte cognitive permet de résoudre la tâche en l'absence d'indices visuels ; dans le test de compétition, les rats avec lesion du fornix sont plus rapides que les deux autres groupes ; dans ce même test, le groupe de rat contrôle peut être subdivisé en deux : ceux qui vont directement à la plate-forme visible, et ceux qui se dirigent d'abord vers l'ancienne position de la plate-forme.

Pour les essais avec plate-forme visible, les trois groupes simulés (complet, sans *taxon* et sans *planification*) apprennent à rejoindre le but (Fig. 3.4, droite). Une différence avec les résultats expérimentaux concerne le groupe *planification* (c'est-à-dire sans *taxon*) : la stratégie de *planification*, ne semble pas pouvoir avoir un niveau d'efficacité aussi élevé que celui du *taxon*, ce qui se traduit par une stagnation autour de 50 pas de temps. Les groupes *taxon* et contrôle ont, eux, des performances similaires. Les essais avec plate-forme cachée donnent bien lieu à une forte dégradation de la performance du groupe *taxon* vis-à-vis de celle des deux autres groupes, sans qu'un apprentissage au cours de ces trois essais ne vienne l'améliorer. Enfin, lors du test de compétition, le groupe *taxon* est, comme dans les données expérimentales, significativement plus rapide que les deux autres. En revanche, le groupe *planification* est aussi significativement plus mauvais que le groupe contrôle.

L'ensemble de ces résultats reproduit nombre de propriétés saillantes de l'expérience de Devan et White (1999), sans pour autant être aussi clairs que ceux obtenus dans l'expérience précédente. Ainsi, lorsque la plate-forme est visible, la tâche est nettement plus facile pour la stratégie *taxon* que dans la tâche de Pearce et al. car l'indice visuel et la plate-forme sont confondus. Cela se traduit par une absence de coopération entre stratégies lors de ces essais : le *taxon* contrôle l'essentiel du comportement (voir Fig. 3.5). Lorsque la plate-forme est dissimulée, il ne peut y avoir coopération, le *taxon* n'ayant aucun indice visuel proche sur lequel s'ancrer. La *planification* assume alors seule le guidage, et la forte proportion d'explo-

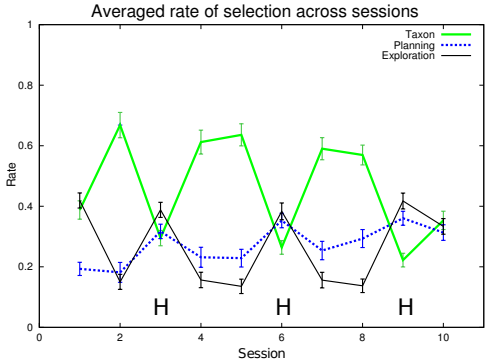

FIG. 3.5 – *Taux de sélection moyen des stratégies au cours des sessions pour le groupe contrôle. H : essais où la plate-forme est cachée (hidden).*

ration en Fig. 3.5 (essais H) confirme que son efficacité est limitée. Une amélioration de notre implémentation de cette stratégie devrait corriger ce défaut. Enfin, durant le test de compétition, il apparaît que le groupe contrôle peut être subdivisé en un groupe répondant au stimulus visuel (59% des individus) et en un autre déviant d'abord sa trajectoire vers l'ancienne position de la plate-forme (41%), ce qui est très similaire aux résultats expérimentaux. Les premiers sélectionnent préférentiellement le *taxon* alors que les seconds favorisent la *planification*.

La différence de performance entre les rats simulés avec *planification* et les rats réels avec lésion du DLS peuvent s'expliquer de deux manières : soit notre implémentation de la stratégie de planification est moins efficace que celle que les rats utilisent, soit une capacité résiduelle de *taxon* fournit le complément d'information nécessaire pour atteindre le niveau de performance des autres groupes. Cette capacité résiduelle peut à la fois être due à des lésions du DLS incomplètes ne désactivant pas totalement le circuit, et à des circuits de *taxon* n'impliquant pas le DLS[2]. Cette deuxième possibilité a l'avantage de pouvoir éventuellement expliquer le fait que les rats avec lésion du DLS sont moins performants aux essais avec plate-forme cachée qu'à ceux avec plate-forme visible. En effet, si les rats DLS utilisent une capacité résiduelle de *taxon* dans les essais à plate-forme visible, lorsqu'elle est cachée, notre modèle prédit que le circuit de sélection aura tendance à continuer de sélectionner cette stratégie devenue inefficace, d'où un niveau de performance moindre que si seul la *planification* était disponible. Enfin, ce *taxon* résiduel expliquerait le niveau de performance similaire des rats contrôles et DLS, là où notre simulation montre que la *planification* seule devrait être significativement moins efficace. Il apparaît nécessaire de concevoir une simulation avec dégradation partielle des capacités de *taxon*, plutôt qu'une suppression totale, afin de vérifier si cette hypothèse peut effectivement expliquer l'ensemble de ces différences.

[2]On pense en particulier au colliculus supérieur : il est en mesure de générer seul des mouvements d'orientation, constituant de base d'une stratégie *taxon*, et son implication dans les stratégies de navigation S-R semble confirmée expérimentalement (Felsen et Mainen, 2008).

3.1.3 Discussion

Au vu des résultats des simulations de ces deux expériences, le modèle proposé semble avoir capturé certains éléments fondamentaux des processus de sélection des stratégies de navigation chez le rat. Cela nous permet de proposer les réponses suivantes aux questions posées : un mécanisme de sélection de stratégies utilisant les mêmes informations d'entrées que les stratégies pour apprendre laquelle de ces stratégies est la plus efficace dans un contexte donné, par un simple apprentissage par renforcement portant sur la valeur de la direction choisie, peut rendre compte des effets de compétition et de coopération entre stratégies. Le seul critère de récompense de cet apprentissage est donc la valeur de récompense prédite pour chaque direction. L'utilisation en entrée des mêmes informations que celles utilisées par l'ensemble des stratégies semble également rendre compte des interactions entre les différents types d'indices sensoriels.

Les différences restantes sont en grande partie dues à une difficulté méthodologique. En effet, les effets exacts des lésions sont difficiles à évaluer et donc difficiles à transcrire dans un modèle : d'une part, une lésion est rarement complète, alors que dans un modèle il est difficile de ne supprimer que partiellement un module ; d'autre part, l'animal peut être en mesure de compenser sa lésion par des processus d'adaptation non-modélisés et difficilement modélisables. Une suite naturelle de cette étude serait donc d'aborder des résultats expérimentaux obtenus avec des méthodes plus ciblées, par exemple les modifications génétiques chez la souris affectant des sites de plasticités très spécifiques ou permettant des inactivations réversibles.

Ce travail apporte également une contribution intéressante du point de vue de l'apprentissage par renforcement. En effet, deux techniques sont utilisées : une sans modèle[3], le *taxon*, et l'autre avec modèle[4], la *planification*. L'apprentissage par ces techniques ne peut converger que grâce à un processus d'exploration permettant de tester l'ensemble des couples état-action (comme cela a été évoqué dans la discussion du précédent chapitre, section 2.4). Cette exploration est en général le résultat d'un processus aléatoire *ad hoc* intégré au modèle, et pour lequel se pose toujours la question de savoir quand il doit être favorisé et sur la base de quel critère (par exemple, en début d'apprentissage, mais aussi lorsqu'un changement survient). Ici, nous avons explicitement ajouté l'exploration comme troisième module, externe aux deux autres, et c'est le système de sélection de stratégie qui, de lui-même, favorise l'exploration lorsqu'aucune des deux autres stratégies ne donne satisfaction (voir par exemple les essais en Fig. 3.5).

Enfin, la stratégie d'apprentissage d'une réponse associée à une activité de cellule de lieux[5] a été utilisée dans de nombreux modèles, en particulier dans celui de Chavarriaga et al. (2005) pour simuler l'expérience de Pearce et al. (1998). Il semble que la *planification* permette de restituer une simulation plus fidèle de cette expérience. Pour autant, les rats semblent en mesure d'utiliser aussi cette stratégie, et elle mériterait sûrement d'être

[3] *model-free*, dans le sens technique utilisé en apprentissage par renforcement.

[4] *model-based*, dans le même contexte.

[5] dite PRTR (*place recognition triggered response*) dans la taxonomie de Trullier et al. (1997), une stratégie S-R utilisant pourtant un élément de la carte cognitive.

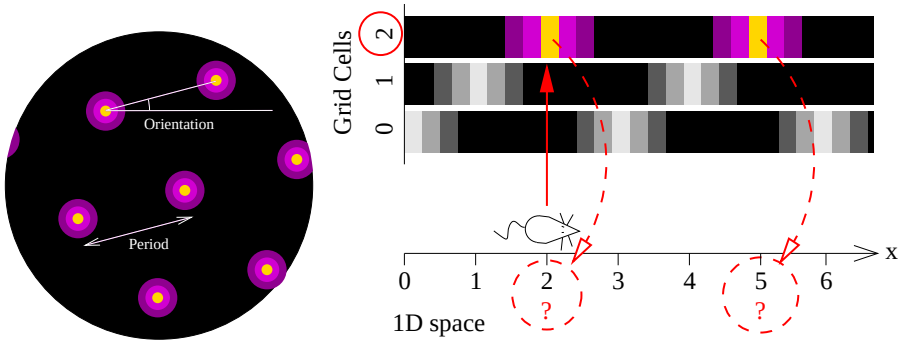

FIG. 3.6 – ***Gauche :*** *Schématisation de l'activité d'une cellule de grille dans l'espace.* ***Droite :*** *Interprétation de l'activité de neurones d'une grille comme un modulo : la position actuelle du rat active un neurone donné (flèche pleine), mais la connaissance de l'activité de ce neurone renseigne sur la position du rat modulo la période de la grille (flèches pointillées).*

ajoutée à notre modèle afin de pouvoir rendre compte d'un certain nombre de comportements relevant de l'habitude.

3.2 Intégration de chemin / retour au point de départ

Publications : (Masson et Girard, 2009)

Comme nous l'avons vu dans la section précédente, les modèles de sélection de stratégies de navigation sont en général limités à un répertoire de stratégies assez limité au regard de la richesse de celles observées chez l'animal. Le modèle de sélection proposé doit être confronté à d'autres protocoles expérimentaux afin de mesurer à quel point il est en mesure d'expliquer les comportements de navigation chez le rongeur. Pour ce faire, nous nous intéressons à la modélisation neuromimétique d'autres stratégies, parmi lesquelles le retour au point de départ (ou *homing*). L'étude menée durant le stage de M2 de Cécile Masson avait plus précisément pour but de démontrer la possibilité de décoder l'activité des cellules de grilles avec un simple réseau de neurones.

En effet, les rongeurs sont capables de revenir directement à leur point de départ après avoir exploré un environnement inconnu, et ce, même en l'absence d'indices allocentriques tels que la vision (Etienne et Jeffery, 2004). Ils y parviennent en intégrant les informations concernant leurs propres déplacements, fournies par le système vestibulaire, la proprioception et la copie efférente des actions, de manière à estimer en continu leur position relativement au point de départ.

Le substrat neural de ce mécanisme d'intégration semble impliquer les cellules de grilles (GC), récemment découvertes (Hafting et al., 2005) dans la bande dorso-latérale du cortex entorhinal médian (dMEC). Ces cellules ont la particularité de décharger suivant un motif constitué de triangles équilatéraux dans le plan de l'espace de locomotion (Fig. 3.6, gauche). Ce motif est caractérisé par sa période (la distance entre deux sommets des triangles) et son orientation (celle d'un des côté des triangles par rapport à une direction absolue). Des cellules voisines dans le dMEC ont des périodes et orientations identiques, mais sont déphasées et semblent donc

appartenir à un groupe de cellules définissant une grille susceptible de couvrir l'ensemble du plan. Des grilles de tailles croissantes sont observées au fur et à mesure que l'on enregistre plus ventralement. Ce motif spatial prend en compte les déplacements propres, puisqu'il est préservé[6] en absence d'indices visuels. Le dMEC est une partie essentielle du mécanisme d'intégration de chemin et de retour au point de départ. Il a été montré que des rats avec une lésion du cortex entorhinal sont incapables d'exhiber le comportement de retour au point de départ (Parron et Save, 2004). De nombreux modèles ont été proposés pour expliquer la formation du motif triangulaire dans les GC ainsi que sa mise à jour lors des déplacement propres (voir McNaughton et al., 2006; Moser et al., 2008, pour des revues), sans cependant expliquer comment l'intégration des mouvements propres dans les GC pouvait ensuite être utilisée pour effectuer un retour au point de départ.

3.2.1 Modèle

Dans un article récent, Fiete et al. (2008) ont proposé une nouvelle façon d'interpréter l'activité des GC : une grille donnée pourrait être vue comme la réalisation neurale d'un calcul de modulo en deux dimensions. Si l'on considère le cas à une dimension, le long de l'un des axes de cette grille, le neurone le plus actif à un moment donné renseigne sur la position de l'animal modulo la période de la grille (Fig. 3.6, droite). Cependant l'information fournie par N grilles de périodes $(\lambda_1, ..., \lambda_N)$ correspond à un encodage de la coordonnée sur l'axe considéré dans le système numéral à base de restes (RNS). Le RNS utilise le théorème chinois des restes (CRT) qui établit qu'à partir d'un ensemble de restes $(r1, ..., r_N)$ et d'un ensemble de nombres premiers entre eux $(\lambda_1, ..., \lambda_N)$ (avec $\Lambda = \prod_{i=1}^{N} \lambda_i$), il existe un unique x modulo Λ tel que $\forall i \in [1, N]$, $x \equiv r_i \pmod{\lambda_i}$. Cela signifie qu'à partir de l'activité[7] de N grilles de périodes λ_i et de même direction, on peut encoder la position de l'animal sur des distances allant jusqu'à Λ (on passe à deux dimensions en considérant deux des trois axes de ces grilles). Ce résultat est généralisable à des périodes qui ne sont pas premières entre elles, auquel cas Λ est le plus petit commun multiple des périodes.

Cette intéressante proposition théorique n'était pas accompagnée d'un modèle computationnel capable d'effectuer ce calcul. Seule était suggérée l'utilisation la méthode de Sun et Yao (1994). Nous avons testé cette méthode et avons pu montrer qu'elle était mathématiquement inadéquate et qu'elle produisait de nombreuses erreurs de décodage.

Nous avons donc proposé un autre modèle de décodage, sur la base de la méthode de reconstruction de x traditionnellement associée au CRT. Considérons les périodes $\hat{\lambda}_i = \frac{\Lambda}{\lambda_i} = \prod_{j \neq i} \lambda_j$: si elles sont premières entre elles, selon le théorème de Bezout, il existe des u_i et des v_i tels que $u_i \lambda_i + v_i \lambda_i = 1$. Si on définit $e_i = v_i \hat{\lambda}_i$, x peut être calculé par la somme pondérée suivante :

$$x = \sum_{i=0}^{N} e_i r_i \tag{3.1}$$

[6] au prix d'une dérive s'accentuant avec le temps, en raison de l'accumulation des erreurs de mesure du mouvement propre par intégration, en l'absence de recalage.

[7] activité qui correspond aux r_i.

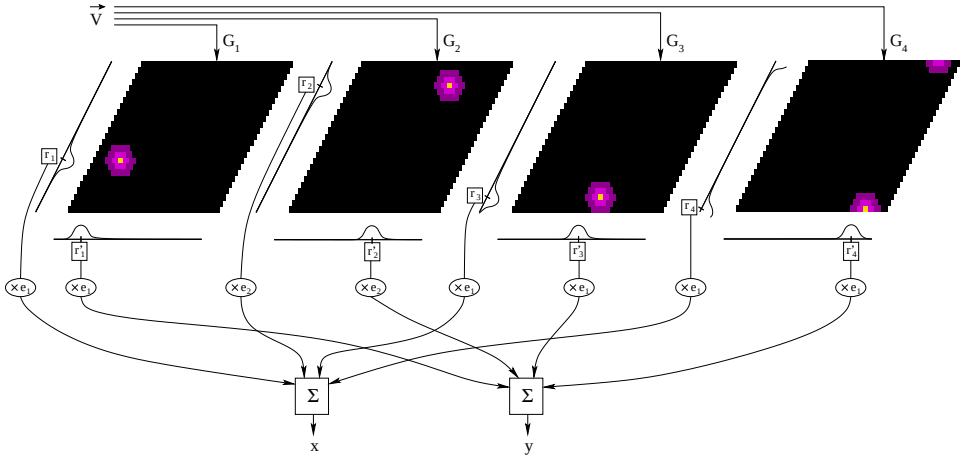

FIG. 3.7 – *Modèle de décodage des cellules de grilles. La vitesse courante $\vec{V}$ permet de mettre à jour l'activité de 4 grilles de périodes différentes ; le calcul des barycentres circulaires de l'activité des grilles projetées sur deux axes permet le décodage explicite des coordonnées du rat (pour des valeurs inférieures à 5km).*

Une solution similaire existe pour le cas où les périodes ne sont pas premières entre elles.

La somme pondérée de l'équation 3.1 peut être calculée par un simple réseau de neurones à une couche, sur la base des r_i. Ces restes correspondent au barycentre de l'activité de chaque grille i projetée sur l'un de ses axes (voir Fig. 3.7). Les grilles sont ici fondées sur le simple modèle de Sheynikhovich (2007). Leur activité est mise à jour à partir de la vitesse propre $\vec{V}$ fournie en entrée. Les grilles utilisées ont des périodes réalistes de $38cm$, $50cm$, $62cm$, et $74cm$, qui permettent théoriquement le décodage de valeurs uniques légèrement supérieures à $5km$. Ces grilles ont des orientations identiques, cependant cette unicité n'est pas un fait expérimental définitivement acquis, encore que certains résultats semblent soutenir cette hypothèse (Barry et al., 2007).

Les simulations réalisées montrent que ce schéma de décodage fonctionne. Les erreurs résiduelles, qui sont causées par la discrétisation des grilles, ont une valeur moyenne de $0.39cm$ (écart-type de 0.19) lorsque le rat simulé parcourt systématiquement une surface de $100 \times 100m$ par pas de $1cm$, une valeur qui semble acceptable au regard de la taille d'un rongeur.

3.2.2 Discussion

Notre modèle a montré l'efficacité de la proposition théorique de Fiete et al. (2008), en proposant une implémentation sous forme de réseau de neurones du décodage des coordonnées d'un rat sur la base de l'activité de ses cellules de grilles. Les valeurs décodées peuvent directement être utilisées comme commande locomotrice par une stratégie de retour au point de départ.

Un modèle computationnel antérieur à la découverte des cellules de grilles (Foster et al., 2000) proposait d'apprendre par renforcement les coordonnées correspondant à chaque cellule de lieux, afin de permettre une

navigation métrique sur la base de ce codage plutôt topologique. Nous avons montré ici que les cellules de grilles, qui fournissent des entrées aux cellules de lieux, sont suffisantes pour reconstruire seules ces coordonnées. Cela signifie que le retour au point de départ est alors possible sans qu'aucune phase d'apprentissage associatif (permettant de stabiliser les cellules de lieux) ne soit nécessaire, ainsi que le démontre l'observation du comportement de retour au point de départ dans des environnements nouveaux et inconnus.

Notre objectif dans cette étude était d'évaluer l'efficacité de notre modèle de décodage, nos simulations ont donc été menées en supposant une absence totale de bruit dans les entrées sensorielles permettant d'évaluer le déplacement propre. Les très faibles erreurs obtenues ne reflètent donc que le bon fonctionnement du décodage. Dans un modèle plus réaliste, les mesures des déplacements propres seraient bruitées et les erreurs s'accumuleraient par l'intégration de ces données par les grilles. Une telle dérive peut être combattue par des recalages de l'activité des grilles par des informations allocentriques qui ne sont pas sujettes à un bruit cumulatif, voir par exemple le modèle de Samu et al. (2009). Le problème du bruit et celui du décodage étant indépendants, notre modèle pourrait fort bien prendre en entrée l'activité de grilles issues d'un modèle intégrant ce recalage.

La possibilité de décoder la position de l'animal ouvre la voie à des stratégies métriques plus complexes que le seul retour au point de départ. En effet, dans notre modèle, nous avons supposé que l'activité des grille était réinitialisée pour chaque changement de point de départ, de sorte que le décodage des grilles produit la position du rat par rapport à ce seul point de départ. L'apprentissage de la position d'autres points d'intérêt permettrait éventuellement la réalisation des calculs vectoriels nécessaires pour rejoindre des buts multiples depuis la position courante (ce que tous les animaux ne semblent pas capables de faire, voir (Etienne et Jeffery, 2004)). De plus, l'existence de cellules encodant les espaces libres et ceux bloqués par des obstacles dans le dMEC (Solstad et al., 2008), là même où l'on trouve aussi des cellules de grilles et des cellules de direction de la tête, permet de spéculer sur l'existence d'un véritable système de planification métrique chez le rat (tel que défini par Trullier et al., 1997). Un modèle d'un tel système, neurobiologiquement plausible et prenant en entrée l'ensemble des informations disponibles dans le dMEC, reste à construire.

3.3 Discussion Générale

Les interactions entre stratégies de navigation concurrentes ou complémentaires semblent être un élément central des capacités d'adaptation des animaux en situation de navigation. Les modèles abordant ces questions sont rares (à notre connaissance, Guazzelli et al. (1998); Girard et al. (2005a); Chavarriaga et al. (2005) et dans une certaine mesure Daw et al. (2005)). Celui de Dollé et al. (2010a) apporte une réponse intéressante d'un point de vue computationnel, dans la mesure où il est le seul qui soit à la fois adaptatif et capable de sélectionner des stratégies fonctionnant sur la base d'algorithmes différents. De surcroît, comme nous l'avons présenté

plus haut, ses capacités à reproduire des résultats expérimentaux sont plutôt prometteuses.

Afin de mesurer à quel point il est généralisable, il semble maintenant indispensable de le confronter à d'autres résultats expérimentaux, et, partant, d'enrichir son répertoire de stratégies. En effet, de nombreuses études expérimentales (par exemple Rondi-Reig et al., 2006) sont susceptibles de mobiliser de nombreuses stratégies, en plus du *taxon* et de la *planification*. C'est dans cet objectif qu'a été mené le travail présenté sur la stratégie de retour au point de départ, et que la question de la construction d'un modèle de navigation métrique plus riche se pose. Pourrait s'y ajouter la modélisation de l'apprentissage d'associations S-R indépendantes de la position du stimulus (par exemple, associer un son à un virage à droite dans un labyrinthe), de la stratégie PRTR, de la stratégie séquentielle égocentrique (apprentissage d'une route, ou séquence d'associations S-R), de la stratégie praxique (apprentissage de séquences strictement motrices), etc.

Enfin, le modèle proposé n'est pas très spécifique quand aux substrats neuronaux de ses modules : par exemple, dans quelle région du cerveau du rat se trouve le module de sélection de stratégie ? Il n'est pas non plus très détaillé dans l'implémentation des modules au substrat bien identifié : l'apprentissage du *taxon* par les circuits dorsaux des BG est effectué par un modèle acteur-critique standard, dans lequel la sélection de l'action résulte d'un simple WTA plutôt que d'un modèle plus fin de la dynamique de sélection, tel qu'un GPR ou un CBG. Ainsi, des travaux préliminaires menés par J. Liénard durant son stage de M2 sur une ancienne version du modèle de sélection de stratégie, avaient montré une amélioration des performances lorsqu'un modèle des BG, avec les effets de persistance évoqués en section 2.1, était utilisé pour les sélections de directions et de stratégies. Des résultats qu'il serait intéressant de reproduire et éventuellement confirmer avec les dernières versions des modèles.

Exécution Motrice

4

Une fois l'action planifiée, sélectionnée, elle doit être transformée en acte moteur. Il apparaît nécessaire d'appréhender la nature de ce processus pour l'implémentation de systèmes neuromimétiques complètement spécifiés, de la perception à l'action, ainsi que pour pouvoir proposer une explication globale, au niveau des interactions entre systèmes, des fonctions motrices.

L'oculomotricité est un modèle mécaniquement simple de motricité, car il n'implique pas une longue chaîne cinématique. Les saccades, en particulier, sont un objet d'étude privilégié en neurosciences, qui a donné lieu à de fructueux aller-retours entre modélisateurs et expérimentateurs depuis plus de 30 ans.

Dans ce cadre particulier, il s'avère de surcroît que le colliculus supérieur n'est pas seulement impliqué dans la mise en œuvre d'actions décidées en amont par le cortex et les ganglions de la base. Il semble qu'il participe activement aux processus perceptifs, mémoriels et de sélection (voir section 1.3.2).

4.1 Transformation spatio-temporelle et géométrie

Publications : (Tabareau et al., 2007)

Ainsi que cela a été mentionné en introduction (section 1.3.2), le colliculus supérieur est constitué d'un empilement de cartes rétinotopiques, où la présence de cibles et les commandes saccadiques sont encodées spatialement. En effet, c'est la position d'une activité de population dans ces cartes qui indique où est le stimulus dans le champ visuel, ou dans quelle direction pointera la prochaine saccade. Chez les espèces étudiées, ces cartes semblent se regrouper en deux grandes familles, celles qui sont linéaires (poisson rouge, souris, rat, etc.) et celles qui sont complexes-logarithmiques (chat, singe, homme). Ces différences peuvent s'expliquer d'un point de vue sensoriel : si l'on considère que la densité de neurones dans les cartes colliculaires est constante, alors, pour conserver toute l'information visuelle, les animaux ayant développé une fovéa doivent déformer la géométrie de leurs carte colliculaires pour lui réserver une plus grande surface, par exemple via une transformation complexe-logarithmique.

Par ailleurs, la commande destinée aux générateurs de saccade est encodée différemment : chaque SBG (vers la droite, la gauche, le haut et le bas) doit recevoir une bouffée d'activité directement proportionnelle à

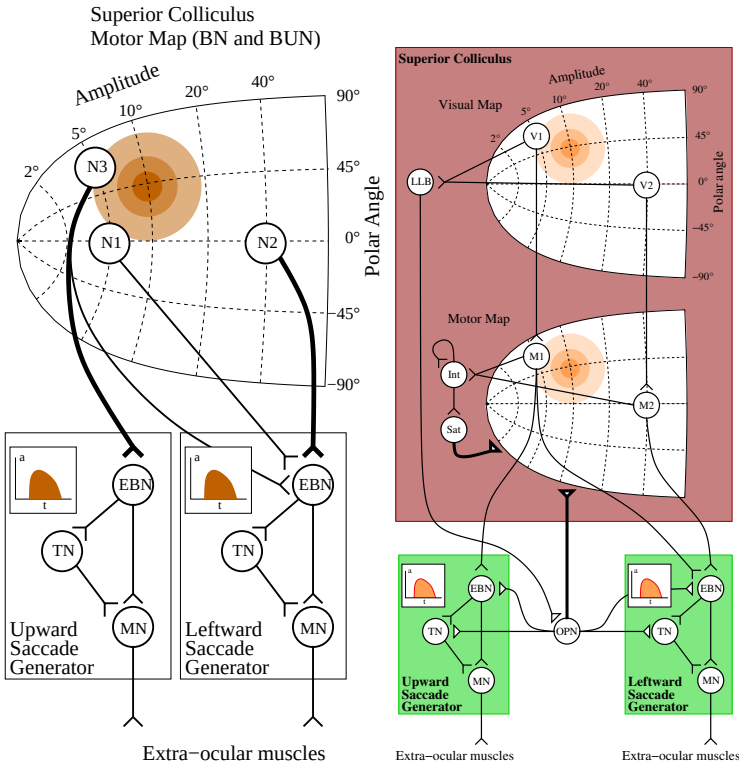

FIG. 4.1 – ***Gauche :*** *Représentation du problème de la transformation spatio-temporelle. La bouffée d'activité stéréotypée dans la carte colliculaire motrice (en haut) doit être transformée en deux bouffées d'activités proportionnelles à l'amplitude des mouvements sur les axes des générateurs de saccades (en bas).* ***Droite :*** *Modèle computationnel de la STT de type "sommation, saturation et inhibition" proposé dans (Tabareau et al., 2007).*

la composante du vecteur mouvement suivant sa direction préférée (voir Fig. 4.1, gauche).

Le processus de transformation de l'encodage spatial dans le SC en un encodage par composante dans les SBG a été étudié depuis longtemps, sous l'appellation de *transformation spatio-temporelle* (ou STT, voir par exemple van Gisbergen et al., 1987), mais il n'a connu des modèles prenant pleinement en compte ses aspects dynamiques (variabilité dans la durée et le profil temorel de l'activité de population) que depuis quelques années (Groh, 2001; Goossens et van Opstal, 2006). Ces modèles proposent que l'activité dans les cartes motrices est transmise aux générateurs de saccade via une somme pondérée. Par exemple, Fig. 4.1 à gauche, les neurones N1 et N2 codent tous les deux pour un mouvement strictement horizontal vers la gauche, et donc ne se projettent que sur le SBG correspondant ; mais le neurone N2, qui code pour une saccade d'amplitude plus importante, a un poids synaptique plus fort (matérialisé par un trait plus épais sur le schéma). Parallèlement, une intégration de l'activité dans la carte, non-pondérée, est réalisée par un circuit parallèle, et lorsque l'intégration atteint un seuil fixe, ce circuit inhibe la carte motrice. Ce mécanisme permet de ne pas recourir à une normalisation incompatible avec les données expérimentales, et donne lieu à des prédictions vérifiées dans Groh (2001) et Goossens et van Opstal (2006).

C'est dans le cadre d'une collaboration avec les Professeurs Alain Ber-

thoz et Daniel Bennequin ainsi qu'avec Nicolas Tabareau que nous avons démontré mathématiquement que la manière dont la STT semble être calculée, sur la base de données expérimentales de la littérature formulées mathématiquement, n'est compatible qu'avec des carte linéaires ou complexes-logarithmiques.

modèle, gluing, etc.

4.1.1 Résultats

Géométrie

Cette démonstration est fondée sur six hypothèses utilisant les notations suivantes : les coordonnées sur la surface colliculaire sont notées $S = X + iY$, celles de la saccade dans le champ visuel $z = \alpha + i\beta$ (ou α est l'azimut et β l'élévation), on s'intéresse à la nature de la bijection $z = \phi(S)$, les coordonnées de la saccade désirée sont notées z_0 et S_0, la consigne envoyée aux SBG (H, horizontal et V, vertical) pour générer la saccade S_0 est $Out_{S_0}(t) = Out^H_{S_0}(t) + iOut^V_{S_0}(t)$.

Si on suppose que :

1. *Somme pondérée* : la sortie du SC à destination des SBG est générée par une somme pondérée de l'activité des cellules saccadiques du colliculus.
$$Out_{S_0}(t) = \int_S w_S \mathcal{A}_{S_0}(S,t)dS \tag{4.1}$$
où $w_S \in \mathbb{C}$ sont les poids de projection du neurone localisé en S vers les générateurs de saccade et où $\mathcal{A}_{S_0}(S,t) \in \mathbb{R}$ est l'activité sur la carte colliculaire en S au temps t pour la génération d'une saccade S_0, supposée à support compact.
2. *Colliculi recollés* : les cartes des deux colliculi sont recollées de manière à former une seule carte sur $\mathbb{R}^2$. Il n'est pas nécessaire de spécifier le mécanisme utilisé pour ce recollement pour la démonstration.
3. *Intégrale invariante* : pour chaque cellule de la carte motrice, le nombre de potentiels d'action émis durant l'exécution de la saccade dépend uniquement de ses coordonnées (X,Y) à la surface du colliculus. Ce qui veut dire que l'on suppose l'existence d'une fonction $K_{\mathcal{A}}$ telle que :
$$\int_t \mathcal{A}_{S_0}(S,t)dt = K_{\mathcal{A}}(S - S_0) \tag{4.2}$$
Cette hypothèse générale englobe le modèle de STT proposé par Groh (2001), repris par Goossens et van Opstal (2006).
4. *Linéarité* : La consigne envoyée du SC vers les SBG est une fonction linéaire de z_0, les coordonnées cartésiennes de la saccade.
$$\int_t Out_{S_0}(t)dt = Cz_0 \quad (C \in \mathbb{R}) \tag{4.3}$$
5. *Carte lisse* : La carte colliculaire est continue et différentiable. $(X,Y) = (0,0)$ correspond à $z = 0$, et les axes horizontaux et verticaux sont alignés avec les axes X et Y en 0 (i.e. la carte est conforme en 0).

6. *Similarité* : Pour toute activité de population respectant l'hypothèse d'intégrale invariante, les poids de projection du SC vers les SBG est une similarité[1] des coordonnées de la saccade exprimées en azimut et élévation. Ce qui veut dire qu'il existe deux complexes a et b tels que :

$$w_S = az + b \tag{4.4}$$

Alors on peut montrer que les seules carte possibles sont linéaires ou complexes-logarithmiques (voir les annexes de l'article, section 6.5 pour le détail de la démonstration) :

$$\frac{X}{B_X} + i\frac{Y}{B_Y} = z \text{ ou } \frac{X}{B_X} + i\frac{Y}{B_Y} = \ln(\frac{z+A}{A}) \tag{4.5}$$

Recollement

Cette preuve supposant l'existence d'un mécanisme de recollement permettant, dans le cas de saccades quasi-verticales impliquant les deux colliculi, la génération de saccades correctes, nous avons également proposé un nouveau schéma de recollement pour les cartes complexes-logarithmiques (voir Fig. 4.2, gauche). En effet, pour les cartes linéaires, il suffit de placer l'activité de population dans chacune des deux cartes, puis de la tronquer pour n'en conserver que la partie correspondant à l'hémichamp visuel de chaque carte pour que le recollement soit correct. En effet, dans ce cas là, la somme des intégrales des deux sous-populations résultantes est bien égale à l'intégrale d'une activité de population dans une seule carte. En revanche, ce n'est plus le cas pour les cartes complexes-logarithmiques : ainsi que l'illustre la figure 4.2 en haut à gauche, la surface hachurée du colliculus gauche n'est pas identique à la partie tronquée (en dehors des hachures) dans le colliculus droit. La génération de saccades suivant ce principe erroné de recollement, similaire à celui proposé par van Gisbergen et al. (1987), conduit à des erreurs systématiques à proximité de la verticale (voir Fig. 4.2, en haut à droite).

Pour résoudre ce problème, nous avons proposé une approche qui consiste à passer progressivement d'une activité de population contenue dans un seul colliculus à une activité partagée entre les deux colliculi, en les modulant par leur proximité au méridien vertical (voir Fig 4.2). Pour ce faire, le stimulus est projeté sur deux cartes d'entrée Inp_L et Inp_R, indépendamment l'une de l'autre, en respectant la géométrie de chaque carte. Ces deux couches projettent neurone à neurone sur les couches motrices, mais sont modulées par la part relative de l'activité contra-latérale contenue dans son hémichamp de prédilection (activité hachurée moins activité totale). Ce mécanisme est implémentable sous forme d'un réseau de neurones et s'est montré capable de corriger les erreurs systématiques de la méthode standard proposée par van Gisbergen et al. (1987) (voir Fig. 4.2, en bas à droite).

[1] une fonction qui préserve les ratios de distances.

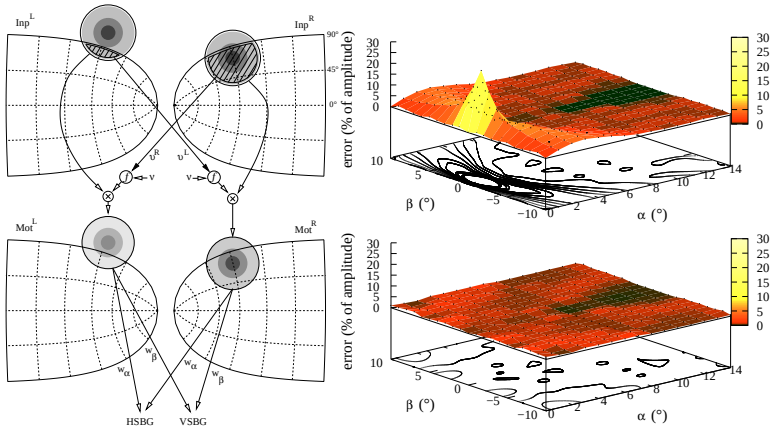

FIG. 4.2 – ***Gauche :*** *modèle de recollement pour les saccades quasi-verticales proposé dans Tabareau et al. (2007).* ***Droite :*** *erreurs (en % de l'amplitude) de saccades générées tous les degrés pour des amplitudes horizontales (α) de 0° à 14° et verticales (β) de -10° à 10° avec les modèles de recollement de van Gisbergen et al. (1987, en haut) et de Tabareau et al. (2007, en bas).*

4.1.2 Discussion

Les hypothèses utilisées pour cette démonstration sont fondées sur des données neurobiologique prises telles quelles ou généralisées avant d'être formulées mathématiquement.

La *somme pondérée* n'est sujette à aucune controverse : le colliculus a des projections directes vers les générateurs de saccade, en particulier les neurones excitateurs à bouffées d'activité (EBN) (Scudder et al., 2002), et ces projections semblent en effet avoir une intensité en relation avec l'amplitude de la saccade représentée (Moschovakis et al., 1998).

Il en est de même des *colliculi recollés* : la capacité des primates à effectuer des saccades verticales correctes démontre que ce recollement est réalisé, on enregistre bien une activité dans les deux colliculi lors des saccades proches de la verticale, et les projection commissurales son probablement le substrat du recollement. Le mécanisme précis du recollement demeure inconnu, mais il n'est point besoin de le préciser pour notre démonstration : sa seule existence suffit.

Plusieurs études récentes sur les singes ont montré que le nombre de potentiels d'action émis par un neurone moteur du colliculus pour une saccade donnée est constant, malgré la présence de perturbations pouvant affecter grandement le décours temporel de l'émission de ces potentiels d'action (Munoz et al., 1996; Soetedjo et al., 2000; Goossens et van Opstal, 2000, 2006; van Opstal et Goossens, 2008). Ces résultats correspondent tout à fait à l'hypothèse d'*intégrale invariante*. Elle reste cependant à vérifier chez les félins dont la morphologie et la physiologie du SC sont trop différente de celles des primates (Grantyn et Moschovakis, 2003) pour que l'on puisse généraliser ces résultats par défaut.

La *linéarité* de la commande reçue par les générateurs de saccade ne fait guère de doute, en particulier au vu des relations observées entre nombre de potentiels d'action et amplitude de la saccade, tant chez le singe que chez le chat, pour les neurones à bouffée du générateur de saccade, excitateurs et inhibiteurs (Keller, 1974; King et Fuchs, 1979; Kaneko et al., 1981;

Yoshida et al., 1982). van Opstal et Goossens (2008) ont récemment mis en évidence que les propriétés non-linéaires des saccades ne sont d'ailleurs pas à rechercher dans les SBG mais plutôt dans les niveaux maximum d'activité dans les cartes motrices du colliculus.

Les propriétés de l'hypothèse de *carte lisse* sont vérifiées pour l'ensemble des cartes colliculaires connues.

Enfin, la *similarité* des poids de projection du colliculus, *exprimés dans l'espace visuel*, est la moins intuitive de nos hypothèses, et s'appuie sur les seuls résultats de Moschovakis et al. (1998) chez le chat, pour les saccades horizontales. Nous avons donc généralisé ce résultat aux saccades verticales, en exigeant une relation de similarité, et nous l'avons étendu aux autres espèces, en particulier le singe. La validité de ces généralisations reste à vérifier expérimentalement. Cependant, sur la base des cinq autres hypothèses et d'une sixième postulant l'existence de cartes linéaires ou complexes-logarithmiques, nous avons pu démontrer cette relation de similarité (voir l'annexe 4.4 de l'article, en section 6.5). Nous avons également étudié la géométrie des cartes colliculaires prédites si cette hypothèse était relaxée en une hypothèse de linéarité des poids (voir l'annexe 4.6 de l'article, en section 6.5).

Le fait que ces propriétés neurobiologiques et la nature des cartes colliculaires puissent être combinées dans ces démonstrations mathématiques renforce leur cohérence et réduit les doutes, discutés ici, quant à leurs incertitudes individuelles (tant celles des données expérimentales, que celles des choix effectués pour leur mise en forme mathématique).

Cette étude laisse entrevoir plusieurs pistes de recherches ultérieures. Tout d'abord, le modèle de réseaux de neurones généré pour simuler la STT et le recollement n'est pas conçu pour sélectionner une cible de saccade lorsque plusieurs sont présentes simultanément (ce qui constitue *a priori* le cas général) : en pareil cas, il effectue une saccade moyenne. Compte-tenu du rôle des ganglions de la base dans les processus de sélection, les boucles tecto-thalamo-baso-tectales sont probablement un substrat sous-cortical de la sélection des cibles des saccades (la boucle CBTC entre la FEF et le circuit oculomoteur des BG constituant, elle, un substrat de plus haut niveau). Un travail préliminaire de modélisation de ces circuits, sur la base des modèles des ganglions de la base et du colliculus supérieur, a été engagé (N'Guyen et al., 2010, , voir section 5.1.3).

D'autre part, l'organisation des neurones du colliculus désignant les cibles des mouvements d'atteinte est mal connue, ainsi que la manière dont ils encodent ces cibles. Suite aux résultats obtenus pour les saccades, il serait intéressant d'explorer, sur la base des données expérimentales, leur éventuelle structuration en cartes : comment la profondeur est-elle prise en compte ? Utilisent-elles elles une géométrie spécifique ? Peut-on tenir à leur égard un raisonnement similaire à celui mené ici ?

Enfin, la coordination des mouvements œil-bras semble pouvoir prendre la forme d'une relation de subsomption dans certain protocoles expérimentaux (Neggers et Bekkering, 2000, 2002). Le découplage et la coordination des cibles des mouvements d'orientation et d'atteinte est un paradigme permettant d'étudier la question non résolue des mécanismes neuronaux de couplage entre les boucles des ganglions de la base (voir section 5.1.1).

Programme de recherche

5

Les travaux présentés sur les thématiques de la sélection de l'action, de la navigation et de l'exécution motrice ne sont pas des contributions isolées les unes des autres. Comme cela a été souligné dans chaque discussion de chapitre, ils se répondent, et les terrains d'interaction encore à explorer définissent pour part mon programme de recherche. Le reste de celui-ci est essentiellement dédié à l'exploration du potentiel des interactions bidirectionnelles entre neurosciences computationnelles et évolution artificielle de réseaux de neurones, amorcée dans le cadre du projet Evo-Neuro.

5.1 Projets Intégratifs

5.1.1 Couplage œil-bras et coordination des circuits des ganglions de la base

J'ai participé avec M. Tran, M. Taïx et Ph. Souères à une première étude abordant la question de la coordination des mouvements d'orientation et de pointage du point de vue des référentiels utilisés, sur une base méthodologique relevant de la robotique humanoïde (Tran et al., 2009a,b). Du point de vue du substrat neural, cette question de coordination permet d'aborder également celle des interactions entre boucles parallèles des ganglions de la base, que j'ai effleurée durant ma thèse (voir Girard et al., 2005a), mais qui a globalement été peu explorée jusqu'ici. En effet, des circuits distincts des ganglions de la base sont dédiés à la sélection des cibles des mouvements des yeux et à ceux du bras. Par ailleurs, le colliculus supérieur, du fait de ses boucles multiples avec les BG et de son implication dans les deux types de mouvements (voir section 4.1) est un substrat subcortical possible de convergence de ces circuits et de coordination de ces mouvements. Nous avons *a priori* à disposition les éléments nécessaires pour bâtir des modèles testant diverses possibilités d'architecture, pour les confronter aux données déjà disponibles sur les situations de couplage et de découplage de ces mouvements, et pour en dériver des prédictions permettant de les distinguer.

5.1.2 Coordination de stratégies de navigation

La poursuite de l'étude des interactions entre stratégies de navigation par la proposition d'un modèle plus proche du substrat neural fait également intervenir l'ensemble des fonctions et régions précédemment présentées. En effet, mener à bien cet objectif nécessitera de faire appel à la modélisation de plusieurs boucles des ganglions de la base, tant dans leurs aspects

de sélection que d'apprentissage par renforcement : une boucle dorso-médiale pour les stratégies S-R, une autre dorso-latérale, dont le rôle exact dans les stratégies utilisant une carte cognitive ou dans la coordination des stratégies reste encore à établir. Seront aussi concernés les circuits colliculaires, pour le taxon egocentré, mais peut-être aussi comme voie de convergence finale déterminant la direction des mouvements de locomotion. Là aussi le savoir-faire déjà acquis sur ces circuits devrait permettre de faire des propositions originales.

5.1.3 Système saccadique

La génération de saccades a l'avantage d'être l'une des fonctions les mieux connues, ayant été (et étant toujours) étudiée avec l'ensemble des méthodes des neurosciences expérimentales et de la psychologie expérimentale, en tant que sujet principal d'investigation ou encore comme moyen d'en étudier d'autres (l'adaptation, la mémoire de travail, l'apprentissage de séquences, l'attention, la fusion multisensorielle, etc.). Cette richesse de données expérimentales a permis de modéliser les circuits saccadiques à tous les niveaux (Girard et Berthoz, 2005), ainsi que de construire des modèles intégratifs, du tronc cérébral au cortex. Ces modèles, au delà de leur seule valeur explicative, permettent d'aborder la question des rôles respectifs des grandes structures du cerveau et de leur structuration en boucles[1].

Les travaux menés sur la conception de modèles des ganglions de la base (Girard et al., 2008), d'une part, et des interactions entre le colliculus supérieur et les générateurs de saccades du tronc cérébral (Tabareau et al., 2007), d'autre part, ont justement pour but l'actualisation du modèle de l'ensemble du circuit saccadique décrit dans la série d'articles de Dominey, Arbib et Schweighofer dans les années 90 (Dominey et Arbib, 1992; Dominey et al., 1995; Arbib et Dominey, 1995; Schweighofer et al., 1996b,a). Cela semble en effet nécessaire puisque, par exemple, les ganglions de la base y sont extrêmement simplifiés, au point qu'ils ne sont pas en mesure d'y sélectionner une cible parmi deux. C'est dans cette optique que nous avons proposé un modèle préliminaire (N'Guyen et al., 2010) intégrant un circuit sous-cortical en charge de la sélection spatiale des cibles et un circuit cortical permettant de moduler cette sélection sur la base de caractéristiques des cibles (ici leur couleur), chacun capable d'apprendre par renforcement, fonctionnant en situation réelle sur le robot-rat *Psikharpax*.

Cette étude se poursuit actuellement par l'ajout de capacités de mémoire de travail et du circuit cortical de sélection spatiale via les champs oculaires frontaux, alors qu'une collaboration avec Q. Wei et D. K. Pai à l'Université de Colombie Britannique a été amorcée pour contrôler leur modèle biomécanique de l'œil (Wei et al., 2010) en lieu et place du modèle simplifié généralement utilisé.

[1] cortico-baso-talamo-corticales, mais aussi tecto-thalamo-baso-tectales, cortico-ponto-cerebello-corticales, etc.

5.2 L'approche EvoNeuro

C'est en collaboration avec le groupe de robotique évolutionniste de l'ISIR (S. Doncieux et J.-B. Mouret) et les équipes de L. Rondi-Reig et A. Arleo (au Laboratoire de Neurobiologie des Processus Adaptatifs) que nous avons initié un nouveau thème de recherche[2], une proposition visant à explorer les interactions entre neurosciences computationnelles et évolution artificielle de réseaux de neurones.

Il s'agit de développer de nouvelles méthodologies à destination à la fois des neurosciences computationnelles (Evo → Neuro) et de l'évolution artificielle de réseaux de neurones (Neuro → Evo). L'évaluation de ces nouvelles méthodologies se fera par leur application à des études de cas, parmi lesquels la sélection de l'action, d'une part, et l'obtention par évolution de circuits cognitifs minimaux, d'autre part.

5.2.1 Les algorithmes évolutionnistes pour modéliser la sélection de l'action

L'utilisation des algorithmes évolutionnistes (AE) dans les neurosciences computationnelles s'est jusqu'ici limité à des tentatives isolées (Arai et al., 1999; Humphries et al., 2005; Keren et al., 2005). Ils y sont cantonnés à un simple ajustement de paramètres à la fin du processus de conception du modèle, lorsque tous les aspects de l'architecture sont fixés, en utilisant des AE basiques. La proposition Evo → Neuro est d'utiliser des AE modernes, fondés sur les résultats récents obtenus dans le domaine, pour aller au delà de la simple optimisation. Ainsi, des AE fondés sur des grammaires génératives permettent d'explorer l'ensemble des solutions possibles à un problème de modélisation, en faisant évoluer la structure même du modèle tout en prenant en compte les contraintes issues des données expérimentales (anatomie, électrophysiologie, comportement).

L'intérêt de cette nouvelle méthodologie de conception de modèle va être d'abord testé dans le cadre de problèmes déjà abordés en neurosciences computationnelles. Cela concerne d'une part les stratégies d'exploration dans le cadre de la navigation chez le rongeur, une tâche assurée par nos partenaires du laboratoire de Neurobiologie des Processus Adaptatifs, que je ne détaillerai pas plus avant ici. D'autre part, il s'agit de revisiter les questions de modélisation des circuits de sélection de l'action, les ganglions de la base, mais aussi la formation réticulée médiale (mRF).

Même si les nombreux modèles des BG tiennent de mieux en mieux compte de l'ensemble de la connectivité de ces circuits (voir Chapitre 2), aucun d'eux ne l'exploite totalement à l'heure actuelle. Notre premier objectif sera donc d'utiliser un AE multi-objectif pour proposer un nouveau modèle des BG plus complet et plus performant que les précédents. Un résultat préliminaire (Liénard et al., 2010) a été obtenu sur ce sujet dans le cadre de la thèse de J. Liénard : les paramètres des modèles GPR et CBG y ont été soumis à une évolution artificielle sur la base de deux objectifs définissant la fonction de WTA.

[2]dans le cadre du projet ANR-09-EMER-005 financé par le programme *Domaines Emergents* de l'Agence Nationale de la Recherche.

La formation réticulée médiale dans le mésencéphale semble être un proto système de sélection de l'action. Elle interagit avec les ganglions de la base, mais a la forme d'un réseau de type small-world. Seuls deux modèles de la mRF ont été jusqu'ici proposés (Kilmer et al., 1969; Humphries et al., 2007), ils sont tous deux basés sur un formalisme ne relevant pas des réseaux de neurones artificiels et proposant des recâblages à chaque pas de temps de simulation, ce qui semble peu crédible. Étant donné les connaissances anatomiques disponibles sur l'organisation générale de la mRF, il semble possible de définir une grammaire générative susceptible d'être utilisée par un AE pour en proposer un modèle réaliste et efficace.

Enfin, les BG et la mRF sont interconnectés via le noyau pédonculopontin, la nature de ces interactions est mal connue et n'a jamais été modélisée, notre objectif à long terme est donc de proposer un ensemble de prédictions sur la question, à tester expérimentalement.

5.2.2 Les neurosciences computationnelles pour enrichir l'évolution artificielle de réseaux de neurones

Les neurosciences computationnelles sont en retour susceptible de faire progresser les AE. En effet, la robotique évolutionniste ambitionne depuis plus de 10 ans de produire des architectures de contrôle robotiques à base de réseaux de neurones obtenues par évolution artificielle et manifestant un certain nombre de fonctions relevant de la cognition (mémoire, navigation, sélection de l'action, apprentissage, etc.). Cet objectif est cependant loin d'être atteint et les résultats obtenus se cantonnent pour la plupart à du contrôle moteur réactif.

On peut noter que l'une des seules métaphores du cerveau importée dans ces méthodes est l'usage de modèles de neurones (et parfois d'oscillateurs assimilables à des abstractions de générateur de rythmes centraux) comme constituants de base. Or la conception de modèles computationnels du cerveau se fonde sur de nombreux autres types de structures. Par exemple les champs de neurones, cartes à deux dimensions dont la dynamique résulte de connexions latérales stéréotypées dépendantes de la distance. De plus, les modèles computationnels des neurosciences incorporent de nombreuses autres connaissances sur le fonctionnement du cerveau. Ainsi, les entrées sensorielles ne sont en général pas encodées par une entrée unique prenant des valeurs sur $\mathbb{R}$, mais par une décomposition du signal sur une population de neurone, chacun ayant son champ récepteur spécialisé sur une plage de données.

La partie Neuro $\rightarrow$ Evo du projet a donc pour objectif de recenser un certain nombre de ces connaissances systématiquement utilisées en neurosciences computationnelles, de les incorporer comme briques de base à disposition des AE et d'évaluer leur efficacité en essayant de générer des circuits exhibant un certain nombre de fonctions de base identifiées dans le cerveau et qui semblent nécessaires à un système cognitif (mémoire de travail, auto-calibration, sélection, etc.).

Là aussi, un premier résultat concernant l'usage de modules aux connections stéréotypées plutôt que de neurones isolés a permis de faire évoluer un réseau aux propriétés de sélection proches de celles des ganglions de la base (Mouret et al., 2010), ce qu'un codage direct classique n'a

pas permis d'obtenir. Parallèlement, P. Tonelli et J.-B. Mouret travaillent à l'évolution de systèmes d'apprentissage par renforcement, et T. Pinville et S. Doncieux à celle d'une mémoire de travail.

Articles

6

6.1 (Girard et al, 2008)

Neural Networks 21 (2008) 628–641

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2008 Special Issue

Where neuroscience and dynamic system theory meet autonomous robotics: A contracting basal ganglia model for action selection

B. Girard [a,*], N. Tabareau [a], Q.C. Pham [a], A. Berthoz [a], J.-J. Slotine [b]

[a] *Laboratoire de Physiologie de la Perception et de l'Action, UMR7152, CNRS - Collège de France, 11 place Marcelin Berthelot, 75231 Paris Cedex 05, France*

[b] *Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

ARTICLE INFO

Article history:
Received 15 March 2007
Received in revised form
7 March 2008
Accepted 7 March 2008

Keywords:
Action selection
Basal ganglia
Computational model
Autonomous robotics
Contraction analysis

ABSTRACT

Action selection, the problem of choosing what to do next, is central to any autonomous agent architecture. We use here a multi-disciplinary approach at the convergence of neuroscience, dynamical system theory and autonomous robotics, in order to propose an efficient action selection mechanism based on a new model of the basal ganglia. We first describe new developments of contraction theory regarding locally projected dynamical systems. We exploit these results to design a stable computational model of the cortico-baso-thalamo-cortical loops. Based on recent anatomical data, we include usually neglected neural projections, which participate in performing accurate selection. Finally, the efficiency of this model as an autonomous robot action selection mechanism is assessed in a standard survival task. The model exhibits valuable dithering avoidance and energy-saving properties, when compared with a simple if-then-else decision rule.

1. Introduction

Action selection is the problem of motor resource allocation an autonomous agent is faced with, when attempting to achieve its long-term objectives. These may vary from survival and reproduction to delivering letters to researchers' offices, depending on the nature of the considered agent (animal, robot, etc.). Action selection is a topic of interest in various disciplines, including ethology, artificial intelligence, psychology, neuroscience, autonomous robotics, etc. We address here the question of action selection for an autonomous robot, using a computational model of brain regions involved in action selection, namely the cortico-baso-thalamo-cortical loops. In order to avoid unwanted dynamical behaviors resulting from a highly recurrent network, we use contraction analysis (Lohmiller & Slotine, 1998) to obtain a rigorous proof of its stability. The efficiency of this action selection mechanism (ASM) is assessed using a standard minimal survival task in a robotic simulation.

The basal ganglia are a set of interconnected subcortical nuclei common to all vertebrates and involved in numerous processes, from motor functions to cognitive ones (Middleton & Strick, 1994; Mink, 1996). Their role is interpreted as a generic selection circuit, and they have been proposed to form the neural substrate of action selection (Krotopov & Etlinger, 1999; Mink, 1996; Redgrave, Prescott, & Gurney, 1999). The basal ganglia are included in cortico-baso-thalamo-cortical loops (Fig. 1), five main loops have been identified in primates (Alexander, Crutcher, & DeLong, 1990; Alexander, DeLong, & Strick, 1986; Kimura & Graybiel, 1995): one motor, one oculomotor, two prefrontal and one limbic loop. Within each of these loops, the basal ganglia circuitry is organized in interacting channels, among which selection occurs. Depending on the considered loop, this selection may concern, for example, the target of an upcoming saccadic movement, the target of a reaching movement or the piece of information to be stored in working memory. The output nuclei of the basal ganglia are inhibitory and tonically active, and thus maintain their targets under sustained inhibition. Selection occurs *via* disinhibition (Chevalier & Deniau, 1990): the removal of the inhibition exerted by one channel on its specific target circuit allows the activation of that circuit. When considering action selection, the basal ganglia channels are thought to be associated to competing action primitives. Given sensory and motivational inputs, the basal ganglia are thus supposed to arbitrate among these actions and to allow the activation of the winner by disinhibiting the corresponding motor circuits.

The considered network contains a large number of closed loops, from the large cortico-baso-thalamo-cortical loop, to small loops formed by the interconnections between nuclei within the basal ganglia and between the thalamus and the cortex. A system with such a structure may exhibit varied dynamical behaviors, some of which should be avoided by an ASM, like reaching a standstill state which does not depend anymore on the

* Corresponding author. Tel.: +33 1 44 27 13 91; fax: +33 1 44 27 13 82.
E-mail address: benoit.girard@college-de-france.fr (B. Girard).

0893-6080/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2008.03.009

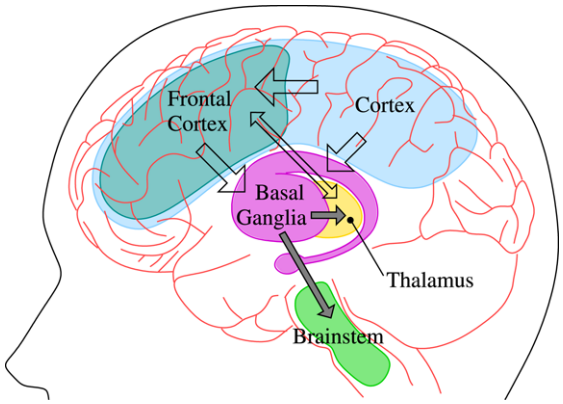

Fig. 1. Cortico-baso-thalamo-cortical loops. The basal ganglia receive inputs from the whole cortex, but establish loops with the frontal areas only. Shaded arrows: inhibitory projections.

external input. This motivates the use of a theoretical framework to study the dynamics of basal ganglia models. We propose to use contraction analysis (Lohmiller & Slotine, 1998) in order to guide the design of a new model of the basal ganglia whose stability can be formally established. Contraction analysis is a theoretical tool used to study the dynamic behavior of nonlinear systems. Contraction properties are preserved through a number of particular combinations, which is useful for a modular design of models.

Numerous computational models of the BG have been proposed in order to investigate the details of the operation of the basal ganglia disinhibition process (see Gillies & Arbruthnott, 2000; Gurney, Prescott, Wickens, & Redgrave, 2004, for recent reviews). Among these, the model proposed by Gurney, Prescott, and Redgrave (2001a, 2001b) (henceforth the GPR model) has been successfully tested as an action selection mechanism for autonomous agents (Girard, Cuzin, Guillot, Gurney, & Prescott, 2003; Girard, Filliat, Meyer, Berthoz, & Guillot, 2005; Montes-Gonzalez, Prescott, Gurney, Humphries, & Redgrave, 2000; Prescott, Montes-Gonzalez, Gurney, Humphries, & Redgrave, 2006). In particular, it was shown to be able to solve a minimal survival task, and, compared with a simpler winner-takes-all mechanism, displayed dithering avoidance and energy-saving capabilities.

We present here an action selection mechanism based on a contracting computational model of the basal ganglia (or CBG). In order to adapt the contraction theory to the analysis of rate-coding artificial neural networks, we first extend it to locally projected dynamical systems (Section 2). Using the resulting neuron model and contraction constraints on the model's parameters, we build a computational model of the basal ganglia including usually neglected neural connections (Section 3). We then check the selection properties of the disembodied model and compare them to those of the GPR, so as to emphasize the consequences of using contraction analysis (Section 4). We finally test its efficiency in a survival task similar to the one used to evaluate the GPR (Girard et al., 2003), and emphasize its dithering avoidance and energy-saving properties by comparing it to a simple if-then-else decision rule (Section 5).

Preliminary versions of the basal ganglia computational model were presented in Girard, Tabareau, Berthoz, and Slotine (2006) and Girard, Tabareau, Slotine, and Berthoz (2005).

2. Nonlinear contraction analysis for rate-coding neural networks

Basically, a nonlinear time-varying dynamic system is said to be *contracting* if initial conditions or temporary disturbances are forgotten exponentially fast, that is, if any perturbed trajectory returns to its nominal behavior with an exponential convergence rate. Contraction is an extension of the well-known *stability* analysis for linear systems. It has the desirable feature of being preserved through hierarchical and particular feedback combinations. Thus, as we will see below, contraction analysis is an appropriate tool to study stability properties of rate-coding neural networks.

In addition, when a system is contracting, it is sufficient to find a particular bounded trajectory to be sure that the system will eventually tend to this trajectory. Thus contraction theory is a convenient way to analyze the dynamic behavior of a system without linearized approximations.

2.1. Contraction theory

We summarize the differential formulation of contraction analysis presented in Lohmiller and Slotine (1998). Contraction analysis is a way to prove the exponential stability of a nonlinear system by studying the properties of its Jacobian. Consider an n-dimensional time-varying system of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ and $t \in \mathbb{R}_+$ and $\mathbf{f}$ is a $n \times 1$ nonlinear vector function which is assumed in the remainder of this paper to be real and smooth, in the sense that all required derivatives exist and are continuous. This equation may also represent the closed-loop dynamics of a neural network model of a brain structure. We recall below the main result of contraction analysis (see Lohmiller and Slotine (1998), for a proof and more details).

Theorem 1. *Consider the continuous-time system (1). If there exists a uniformly positive definite metric*

$$\mathbf{M}(\mathbf{x}, t) = \mathbf{\Theta}(\mathbf{x}, t)^{\mathrm{T}}\mathbf{\Theta}(\mathbf{x}, t)$$

such that the generalized Jacobian

$$\mathbf{F} = (\dot{\mathbf{\Theta}} + \mathbf{\Theta}\mathbf{J})\mathbf{\Theta}^{-1}$$

is uniformly negative definite, then all system trajectories converge exponentially to a single trajectory with convergence rate $|\lambda_{\max}|$, where $\lambda_{\max}$ is the largest eigenvalue of the symmetric part of $\mathbf{F}$.

The symmetric part of a matrix $\mathbf{A}$ is $\mathbf{A}_s = 1/2(\mathbf{A} + \mathbf{A}^{\mathrm{T}})$. A matrix $\mathbf{A}(\mathbf{x}, t)$ is uniformly positive definite if there exists $\beta > 0$ such that

$$\forall \mathbf{x}, t \quad \lambda_{\min}(\mathbf{A}(\mathbf{x}, t)) \geq \beta.$$

2.2. Neural networks and locally projected dynamical systems

Networks of leaky integrators are widely used to model the behavior of neuronal assemblies (Dayan & Abbott, 2001). A leaky-integrator network is usually described by the following set of equations

$$\tau_i \dot{x}_i = -x_i(t) + \sum_{j \neq i} K_{ji} x_j(t) + I(t)$$

where $x(t)$ is the synaptic current of a neuron, τ_i its time constant, K_{ji} the synaptic projection weight from neuron j to neuron i and $I(t)$ the input coming from an external source. Next, $x(t)$ is converted into a non-negative firing rate $y(t)$ using a transfer function, for instance

$$y(t) = \max(x(t), 0) = [x(t)]_+.$$

Another way to enforce non-negativity of the firing rate is to use through *locally projected dynamical systems* (lPDS in short). These systems were introduced in Dupuis and Nagurney (1993)

and further analyzed in Zhang and Nagurney (1995). Related ideas can be found in the standard parameter projection method in adaptive control (Ioannou & Sun, 1996; Slotine & Coetsee, 1986). A lPDS is given by

$$\dot{\mathbf{x}} = \mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{f}(\mathbf{x}, t)) \tag{2}$$

where $\mathbf{\Omega}$ is a convex subset of the state space and $\mathbf{\Pi}_{\mathbf{\Omega}}$ is the vector-projection operator on $\mathbf{\Omega}$ given by

$$\mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{v}) = \lim_{h \to 0^+} \frac{\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v}) - \mathbf{x}}{h}.$$

In the above equation, $\mathbf{P}_{\mathbf{\Omega}}$ denotes the point-projection operator on the convex $\mathbf{\Omega}$ defined as

$$\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{y} \in \mathbf{\Omega}} \|\mathbf{x} - \mathbf{y}\|.$$

Intuitively, if $\mathbf{x}$ is in the interior of $\mathbf{\Omega}$ then $\mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{v}) = \mathbf{v}$. If $\mathbf{x}$ is on the boundary of $\mathbf{\Omega}$, then $\mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{v})$ is the maximal component of $\mathbf{v}$ that allows the system to remain within $\mathbf{\Omega}$. In particular, it is easy to see that any trajectory starting in $\mathbf{\Omega}$ remains in $\mathbf{\Omega}$.

Note that Eq. (2) does not define a classical ordinary differential equation since its right-hand side can be discontinuous due to the projection operator. However, under some conditions on $\mathbf{f}$ and $\mathbf{\Omega}$ (similar to the Cauchy–Lipschitz conditions for classical ordinary differential equations, see Dupuis and Nagurney (1993) and Filippov (1988) for more details), existence, uniqueness and some qualitative properties can be established for the solutions of (2). For our purpose, we recall here that any solution $\mathbf{x}$ of (2) is continuous and right differentiable for all t. In the remainder of this article, we make the additional assumption that the set of time instants when $\mathbf{x}(t)$ is not differentiable has measure zero.

Within the above framework, the dynamics of a neural network can now be given in the matrix form as

$$\dot{\mathbf{x}} = \mathbf{\Pi}_{\mathbb{H}_n}(\mathbf{x}, \mathbf{W}\mathbf{x} + \mathbf{I}(t)) \tag{3}$$

where $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))^{\mathrm{T}}$ is the states of the neurons, $\mathbf{W}$ is the $n \times n$ matrix whose diagonal elements represent the leaking rate of the neurons and whose non-diagonal elements represent the synaptic projection weight, $\mathbf{I}(t)$ is the vector of external inputs. Finally, $\mathbb{H}_n$ is a *regular n-cube* defined as follows

Definition 1. A *regular n-cube* $\mathbb{H}_n$ is a subset of $\mathbb{R}^n$ defined by

$$\mathbb{H}_n = \{(x_1, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n : \forall i,\ m_i \le x_i \le M_i\}$$

where $m_1, \ldots, m_n, M_1, \ldots, M_n \in \mathbb{R}$.

Intuitively, a regular n-cube is an n-cube whose edges are parallel to the axes.

In practice, networks of leaky integrators described by lPDS as above and their classical counterparts with transfer functions show very similar behavior. However, the stability properties of lPDS networks can be rigorously established through contraction theory (see the next section), which makes them interesting from a theoretical viewpoint.

2.3. Contraction analysis of locally projected dynamical system on regular n-cubes

Contraction analysis for systems subject to convex constraints has already been discussed in Lohmiller and Slotine (2000). However, in that work, the projection applied to constrain the system in the convex region depends on the metric which makes the original system contracting. Thus, we cannot use this result here since our projection operator must not depend on the neural network

Since the contraction condition is local, a lPDS can only be contracting if the original, un-projected, system is contracting within $\mathbf{\Omega}$. The converse implication is not true in general, because the projection operator can deeply modify the system's behavior along the boundary of $\mathbf{\Omega}$. We now introduce some definitions in order to be able to state this converse implication in some particular cases.

Definition 2. Let $\mathbf{x} \in \delta\mathbf{\Omega}$ where $\delta\mathbf{\Omega}$ denotes the boundary of $\mathbf{\Omega}$. The set of *inward normals* to $\mathbf{\Omega}$ at $\mathbf{x}$ is defined as

$$N_{\mathbf{\Omega}}(\mathbf{x}) = \{\mathbf{n} : \forall \mathbf{y} \in \mathbf{\Omega}, \mathbf{n}^{\mathrm{T}}(\mathbf{x} - \mathbf{y}) \le 0\}.$$

If $\mathbf{x} \in \mathbf{\Omega} - \delta\mathbf{\Omega}$ then we set $N_{\mathbf{\Omega}}(\mathbf{x}) = \{\mathbf{0}\}$.

Definition 3. A metric $\mathbf{M}$ is said to be *compatible* with a convex set $\mathbf{\Omega}$ if there exists a coordinate transform $\mathbf{\Theta}$ such that $\mathbf{\Theta}^{\mathrm{T}}\mathbf{\Theta} = \mathbf{M}$ and

$$\forall \mathbf{x} \in \delta\mathbf{\Omega}, \forall \mathbf{n} \in N_{\mathbf{\Omega}}(\mathbf{x}), \quad \mathbf{\Theta}\mathbf{n} \in N_{\mathbf{\Theta}\mathbf{\Omega}}(\mathbf{\Theta}\mathbf{x}).$$

In this case, we say that $\mathbf{\Theta}$ is a square root of $\mathbf{M}$ which is *compatible* with $\mathbf{\Omega}$.

We can give a simple sufficient condition for a metric to be compatible with a regular n-cube.

Proposition 1. *Any diagonal positive definite metric* $\mathbf{M}$ *is compatible with any regular n-cube* $\mathbb{H}_n$.

Proof. Let $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}} \in \delta\mathbb{H}_n$. An inward normal $\mathbf{n} = (n_1, \ldots, n_n)^{\mathrm{T}}$ to $\mathbb{H}_n$ at $\mathbf{x}$ is characterized by

$$\begin{cases} n_i \ge 0 & \text{if } x_i = m_i \\ n_i \le 0 & \text{if } x_i = M_i \\ n_i = 0 & \text{if } m_i < x_i < M_i. \end{cases}$$

Since $\mathbf{M}$ is diagonal and positive definite, one has $\mathbf{M} = \operatorname{diag}(d_1^2, \ldots, d_n^2)$ with $d_i > 0$. Consider the coordinate transform $\mathbf{\Theta} = \operatorname{diag}(d_1, \ldots, d_n)$. Clearly, $\mathbf{\Theta}^{\mathrm{T}}\mathbf{\Theta} = \mathbf{M}$ and $\mathbf{\Theta}\mathbb{H}_n$ is a regular n-cube with minimal values $d_1 m_1, \ldots, d_n m_n$ and maximal values $d_1 M_1, \ldots, d_n M_n$. It follows from the characterization above that $\mathbf{\Theta}\mathbf{n} = (d_1 n_1, \ldots, d_n n_n)^{\mathrm{T}} \in N_{\mathbf{\Theta}\mathbb{H}_n}(\mathbf{\Theta}\mathbf{x})$. □

We also need another elementary result.

Lemma 1. *Let* $\mathbf{x} \in \mathbf{\Omega}$ *and* $\mathbf{v} \in \mathbb{R}^n$. *There exists* $\mathbf{n}(\mathbf{x}, \mathbf{v}) \in N_{\mathbf{\Omega}}(\mathbf{x})$ *such that*

$$\mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{v}) = \mathbf{v} + \mathbf{n}(\mathbf{x}, \mathbf{v}).$$

Proof. Let $\mathbf{y} \in \mathbf{\Omega}$. We need to show that $A_{\mathbf{y}} = (\mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{v}) - \mathbf{v})^{\mathrm{T}}(\mathbf{x} - \mathbf{y}) \le 0$. By the definition of $\mathbf{\Pi}_{\mathbf{\Omega}}$, one has

$$A_{\mathbf{y}} = \lim_{h \to 0^+} \frac{1}{h}(\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v}) - (\mathbf{x} + h\mathbf{v}))^{\mathrm{T}}(\mathbf{x} - \mathbf{y}).$$

Next, introduce the terms $\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v})$ and $h\mathbf{v}$ into $(\mathbf{x} - \mathbf{y})$

$$\begin{aligned} A_{\mathbf{y}} = \lim_{h \to 0^+} \frac{1}{h} [&(\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v}) - (\mathbf{x} + h\mathbf{v}))^{\mathrm{T}}(\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v}) - \mathbf{y}) \\ &+ (\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v}) - (\mathbf{x} + h\mathbf{v}))^{\mathrm{T}}(\mathbf{x} + h\mathbf{v} - \mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v})) \\ &+ (\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v}) - (\mathbf{x} + h\mathbf{v}))^{\mathrm{T}}(-h\mathbf{v})]. \end{aligned}$$

The first term in the above equation is non-positive by the property of the point-projection operator. The second term is the negative of a distance and thus is also non-positive. As for the third term, observe that

$$\lim_{h \to 0^+} (\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x} + h\mathbf{v}) - (\mathbf{x} + h\mathbf{v}))^{\mathrm{T}}\mathbf{v} = (\mathbf{P}_{\mathbf{\Omega}}(\mathbf{x}) - \mathbf{x})^{\mathrm{T}}\mathbf{v} = 0$$

since $\mathbf{x} \in \mathbf{\Omega}$. □

We can now state the following theorem

Theorem 2. *Let* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$ *be a dynamical system which is contracting in a constant metric* $\mathbf{M}$ *compatible with a convex set* $\mathbf{\Omega}$*. Then the lPDS* $\dot{\mathbf{x}} = \mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{f}(\mathbf{x}, t))$ *is also contracting in the same metric and with the same contraction rate.*

Proof. Let $\mathbf{\Theta}$ be a square root of $\mathbf{M}$ compatible with $\mathbf{\Omega}$. Consider $\mathbf{z} = \mathbf{\Theta}\mathbf{x}$. By Lemma 1, the system $\mathbf{z}$ is described by

$$\dot{\mathbf{z}} = \mathbf{\Theta}\mathbf{\Pi}_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{F}(\mathbf{z}) + \mathbf{\Theta}\mathbf{n}(\mathbf{x}, \mathbf{f}(\mathbf{x})) \tag{4}$$

where $\mathbf{F}(\mathbf{z}) = \mathbf{\Theta}\mathbf{f}(\mathbf{\Theta}^{-1}\mathbf{z})$.

Consider two particular trajectories of (4) $\mathbf{z}_1$ and $\mathbf{z}_2$. Denote by Δ the squared distance between $\mathbf{z}_1$ and $\mathbf{z}_2$

$$\Delta(t) = \|\mathbf{z}_1(t) - \mathbf{z}_2(t)\|^2 = (\mathbf{z}_1(t) - \mathbf{z}_2(t))^{\mathrm{T}}(\mathbf{z}_1(t) - \mathbf{z}_2(t)).$$

When Δ is differentiable, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Delta &= 2(\mathbf{z}_1 - \mathbf{z}_2)^{\mathrm{T}}(\dot{\mathbf{z}}_1 - \dot{\mathbf{z}}_2)\\ &= 2(\mathbf{z}_1 - \mathbf{z}_2)^{\mathrm{T}}(\mathbf{F}(\mathbf{z}_1) + \mathbf{\Theta}\mathbf{n}(\mathbf{x}_1, \mathbf{f}(\mathbf{x}_1)) - (\mathbf{F}(\mathbf{z}_2)\\ &\quad + \mathbf{\Theta}\mathbf{n}(\mathbf{x}_2, \mathbf{f}(\mathbf{x}_2)))).\end{aligned}$$

Since the metric is compatible with $\mathbf{\Omega}$, $\mathbf{\Theta}\mathbf{n}(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i)) \in N_{\mathbf{\Theta}\mathbf{\Omega}}(\mathbf{z}_i)$ for $i = 1, 2$. Next, by the definition of inward normals, we have $(\mathbf{z}_1 - \mathbf{z}_2)^{\mathrm{T}}\mathbf{\Theta}\mathbf{n}(\mathbf{x}_1, \mathbf{f}(\mathbf{x}_1)) \leq 0$ and $-(\mathbf{z}_1 - \mathbf{z}_2)^{\mathrm{T}}\mathbf{\Theta}\mathbf{n}(\mathbf{x}_2, \mathbf{f}(\mathbf{x}_2)) \leq 0$, from which we deduce

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Delta &\leq 2(\mathbf{z}_1 - \mathbf{z}_2)^{\mathrm{T}}(\mathbf{F}(\mathbf{z}_1) - \mathbf{F}(\mathbf{z}_2))\\ &\leq -2\lambda\Delta(t)\end{aligned}$$

where $\lambda > 0$ is the contraction rate of $\mathbf{f}$ in the metric $\mathbf{M}$.

Since the set of time instants when $\Delta(t)$ is not differentiable has measure zero (see Section 2.2), one has

$$\forall t \geq 0, \quad \Delta(t) = \int_0^t \left(\frac{\mathrm{d}}{\mathrm{d}t}\Delta\right)\mathrm{d}t \leq -2\lambda\int_0^t \Delta(s)\mathrm{d}s$$

which yields by Grönwall's lemma

$$\forall t \geq 0, \quad \Delta(t) \leq \Delta(0)\mathrm{e}^{-2\lambda t}$$

i.e.

$$\forall t \geq 0, \quad \|\mathbf{z}_1(t) - \mathbf{z}_2(t)\| \leq \|\mathbf{z}_1(0) - \mathbf{z}_2(0)\|\mathrm{e}^{-\lambda t}. \quad \square$$

2.4. Combination of contracting systems

One of our motivations for using contraction theory is that contraction properties are preserved under suitable combinations (Lohmiller & Slotine, 1998). This allows both stable aggregation of contracting systems, and variation or optimization of individual subsystems while preserving overall functionality (Slotine & Lohmiller, 2001). We present here three standard combinations of contracting systems which preserve both contraction of the system and diagonality of the metric. Then, constructing our neural network as a lPDS using only those three combinations will give rise to a contracting system in a diagonal metric.

2.4.1. Negative feedback combination

Consider two coupled systems

$$\dot{\mathbf{x}}_1 = \mathbf{f}_1(\mathbf{x}_1, \mathbf{x}_2, t)$$
$$\dot{\mathbf{x}}_2 = \mathbf{f}_2(\mathbf{x}_1, \mathbf{x}_2, t).$$

Assume that system i ($i = 1, 2$) is contracting with respect to $\mathbf{M}_i = \mathbf{\Theta}_i^{\mathrm{T}}\mathbf{\Theta}_i$, with rate λ_i. Assume furthermore that the two systems are connected by *negative feedback* (Tabareau & Slotine, 2006). More precisely, the Jacobian matrices of the couplings verify

$$\mathbf{\Theta}_1\mathbf{J}_{12}\mathbf{\Theta}_2^{-1} = -k\mathbf{\Theta}_2\mathbf{J}_{21}^{\mathrm{T}}\mathbf{\Theta}_1^{-1}$$

with k a positive constant. Hence, the Jacobian matrix of the unperturbed global system is given by

$$\mathbf{J} = \begin{pmatrix}\mathbf{J}_1 & -k\mathbf{\Theta}_1^{-1}\mathbf{\Theta}_2\mathbf{J}_{21}^{\mathrm{T}}\mathbf{\Theta}_1^{-1}\mathbf{\Theta}_2\\ \mathbf{J}_{21} & \mathbf{J}_2\end{pmatrix}.$$

Consider the coordinate transform

$$\mathbf{\Theta} = \begin{pmatrix}\mathbf{\Theta}_1 & \mathbf{0}\\ \mathbf{0} & \sqrt{k}\mathbf{\Theta}_2\end{pmatrix}$$

associated with the metric $\mathbf{M} = \mathbf{\Theta}^{\mathrm{T}}\mathbf{\Theta} > \mathbf{0}$. After some calculations, one has

$$\begin{aligned}\left(\mathbf{\Theta}\mathbf{J}\mathbf{\Theta}^{-1}\right)_s &= \begin{pmatrix}\left(\mathbf{\Theta}_1\mathbf{J}_1\mathbf{\Theta}_1^{-1}\right)_s & \mathbf{0}\\ \mathbf{0} & \left(\mathbf{\Theta}_2\mathbf{J}_2\mathbf{\Theta}_2^{-1}\right)_s\end{pmatrix}\\ &\leq \max(-\lambda_1, -\lambda_2)\mathbf{I}.\end{aligned} \tag{5}$$

The augmented system is thus contracting with respect to the metric $\mathbf{M}$, with rate $\min(\lambda_1, \lambda_2)$.

2.4.2. Hierarchical combination

We first recall a standard result in matrix analysis (Horn & Johnson, 1985). Let $\mathbf{A}$ be symmetric matrix in the form

$$\mathbf{A} = \begin{pmatrix}\mathbf{A}_1 & \mathbf{A}_{21}^{\mathrm{T}}\\ \mathbf{A}_{21} & \mathbf{A}_2\end{pmatrix}.$$

Assume that $\mathbf{A}_1$ and $\mathbf{A}_2$ are positive definite. Then $\mathbf{A}$ is positive definite if

$$\sigma^2(\mathbf{A}_{21}) < \lambda_{\min}(\mathbf{A}_1)\lambda_{\min}(\mathbf{A}_2)$$

where $\sigma(\mathbf{A}_{21})$ denotes the largest singular value of $\mathbf{A}_{21}$. In this case, the smallest eigenvalue of $\mathbf{A}$ satisfies

$$\begin{aligned}\lambda_{\min}(\mathbf{A}) \geq\ & \frac{\lambda_{\min}(\mathbf{A}_1) + \lambda_{\min}(\mathbf{A}_2)}{2}\\ & - \sqrt{\left(\frac{\lambda_{\min}(\mathbf{A}_1) - \lambda_{\min}(\mathbf{A}_2)}{2}\right)^2 + \sigma^2(\mathbf{A}_{21})}.\end{aligned}$$

Consider now the same set-up as in Section 2.4.1, except that the connection is now *hierarchical* and upper bounded. More precisely, the Jacobians of the couplings verify

$$\mathbf{J}_{12} = \mathbf{0}, \quad \sigma^2(\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1}) \leq K.$$

Hence, the Jacobian matrix of the augmented system is given by

$$\mathbf{J} = \begin{pmatrix}\mathbf{J}_1 & \mathbf{0}\\ \mathbf{J}_{21} & \mathbf{J}_2\end{pmatrix}.$$

Consider the coordinate transform

$$\mathbf{\Theta}_\epsilon = \begin{pmatrix}\mathbf{\Theta}_1 & \mathbf{0}\\ \mathbf{0} & \epsilon\mathbf{\Theta}_2\end{pmatrix}$$

associated with the metric $\mathbf{M}_\epsilon = \mathbf{\Theta}_\epsilon^{\mathrm{T}}\mathbf{\Theta}_\epsilon > \mathbf{0}$. After some calculations, one has

$$\left(\mathbf{\Theta}\mathbf{J}\mathbf{\Theta}^{-1}\right)_s = \begin{pmatrix}\left(\mathbf{\Theta}_1\mathbf{J}_1\mathbf{\Theta}_1^{-1}\right)_s & \frac{1}{2}\epsilon(\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1})^{\mathrm{T}}\\ \frac{1}{2}\epsilon\mathbf{\Theta}_2\mathbf{J}_{21}\mathbf{\Theta}_1^{-1} & \left(\mathbf{\Theta}_2\mathbf{J}_2\mathbf{\Theta}_2^{-1}\right)_s\end{pmatrix}.$$

Set now $\epsilon = \sqrt{\frac{2\lambda_1\lambda_2}{K}}$. The augmented system is then contracting with respect to the metric $\mathbf{M}_\epsilon$, with rate λ verifying

$$\lambda \geq \frac{1}{2}\left(\lambda_1 + \lambda_2 - \sqrt{\lambda_1^2 + \lambda_2^2}\right).$$

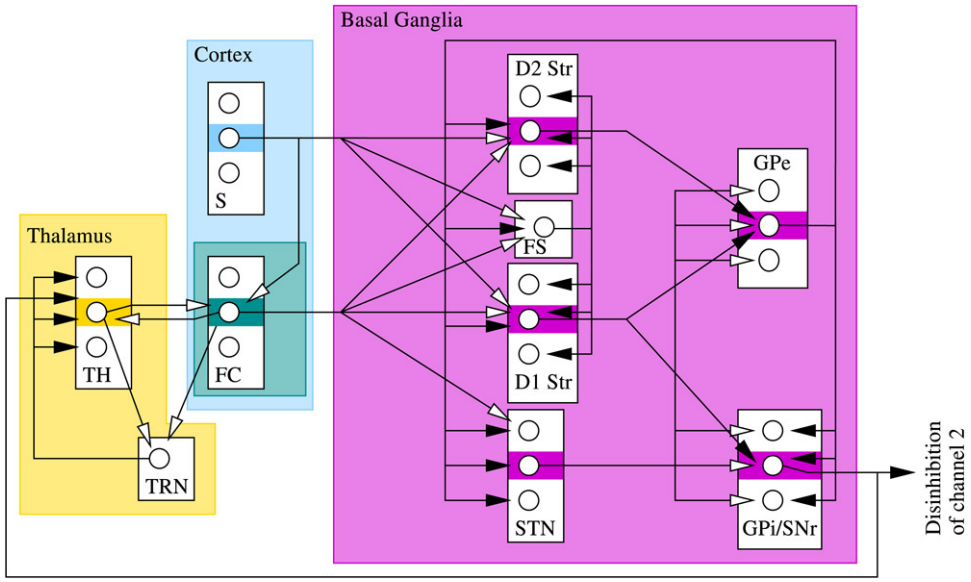

Fig. 2. Basal ganglia model. Nuclei are represented by boxes, each circle in these nuclei represents an artificial rate-coding neuron. In this diagram, three channels are competing for selection, represented by the three neurons in each nucleus. The second channel is represented by colored shading. For clarity, the projections from the second channel neurons only are represented, they are identical for the other channels. White arrowheads represent excitations and black arrowheads, inhibitions. D1 and D2: neurons of the striatum with two respective types of dopamine receptors; STN: subthalamic nucleus; GPe: external segment of the globus pallidus; GPi/SNr: internal segment of the globus pallidus and substantia nigra pars reticulata.

2.4.3. Small gains

In this section, we require no specific assumption on the form of the couplings

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_2 \end{pmatrix}.$$

As for negative feedback, consider the coordinate transform

$$\mathbf{\Theta}_k = \begin{pmatrix} \mathbf{\Theta}_1 & \mathbf{0} \\ \mathbf{0} & \sqrt{k}\mathbf{\Theta}_2 \end{pmatrix} \quad k > 0$$

associated with the metric $\mathbf{M}_k = \mathbf{\Theta}_k^{\mathrm{T}}\mathbf{\Theta}_k > \mathbf{0}$. After some calculations, one has

$$\left(\mathbf{\Theta}_k \mathbf{J} \mathbf{\Theta}_k^{-1}\right)_s = \begin{pmatrix} \left(\mathbf{\Theta}_1 \mathbf{J}_1 \mathbf{\Theta}_1^{-1}\right)_s & \mathbf{A}_k^{\mathrm{T}} \\ \mathbf{A}_k & \left(\mathbf{\Theta}_2 \mathbf{J}_2 \mathbf{\Theta}_2^{-1}\right)_s \end{pmatrix}$$

where $\mathbf{A}_k = \frac{1}{2}\left(\sqrt{k}\mathbf{\Theta}_2 \mathbf{J}_{21} \mathbf{\Theta}_1^{-1} + \frac{1}{\sqrt{k}}\left(\mathbf{\Theta}_1 \mathbf{J}_{12} \mathbf{\Theta}_2^{-1}\right)^{\mathrm{T}}\right)$. Following the result stated at the beginning of Section 2.4.2, if

$$\min_k \sigma^2(\mathbf{A}_k) < \lambda_1 \lambda_2$$

then the augmented system is contracting with respect to the metric $\mathbf{M}_k$ for some k, with rate λ verifying

$$\lambda \geq \frac{\lambda_1 + \lambda_2}{2} - \sqrt{\left(\frac{\lambda_1 - \lambda_2}{2}\right)^2 + \min_k \sigma^2(\mathbf{A}_k)}.$$

3. Model description

Rather than using standard leaky-integrator rate-coding neurons, we use the very similar local projected dynamical system model defined by Eq. (3), where each component of the state vector $\mathbf{x}$ is an artificial rate-coding neuron representing the discharge rate of populations of real neurons. Each competing BG channel in each nucleus is represented by one such neuron, and the corresponding thalamic nucleus and cortical areas are also subdivided into identical channels (Fig. 2). The convergence of cortical sensory inputs on the striatum channels is encoded, for simplicity, by a vector of saliences (one salience per channel). Each salience represents the propensity of its corresponding channel to be selected. Each behavior in competition is associated to a specific channel and can be executed if and only if its level of inhibition decreases below a the inhibition level at rest $y^{\mathrm{GPi}}_{\mathrm{Rest}}$ (ie. the SNr/GPi output when the salience vector is null).

The main difference of our architecture with the recent GPR proposal (Gurney et al., 2001a) is the nuclei targeted by the external part of the globus pallidus (GPe) and the nature of these projections. In our model, the GPe projects to the subthalamic nucleus (STN), the internal part of the globus pallidus (GPi) and the substantia nigra pars reticulata (SNr), as well as to the striatum, as documented in Bevan, Booth, Eaton, and Bolam (1998), Kita, Tokuno, and Nambu (1999) and Staines, Atmadja, and Fibiger (1981). Moreover, the striatal terminals target the dendritic trees, while pallidal, nigral and subthalamic terminals form perineuronal nets around the soma of the targeted neurons (Sato, Lavallee, Lévesque, & Parent, 2000). This specific organization allows GPe neurons to influence large sets of neurons in GPi, SNr and STN (Parent et al., 2000), thus the sum of the activity of all GPe channels influences the activity of STN and GPi/SNr neurons (Eqs. (9) and (11)), while there is a simple channel-to-channel projection to the striatum Eqs. (6) and (7).

The striatum is one of the two input nuclei of the BG. It is mainly composed of GABAergic (inhibitory) medium spiny neurons (MSN). As in the GPR model, we distinguish among them, those with D1 and D2 dopamine receptors and modulate the input generated in the dendritic tree by the dopamine level γ, which here encompasses salience, frontal cortex feedback and GPe projections.

Using the formulation of Eq. (3), the ith neurons ($i \in [1, N]$, with N the number of channels) of the D1 and D2 subparts of the striatum are defined as follows

$$(\mathbf{Wx} + \mathbf{I}(t))_{\mathrm{D1}_i} = \frac{1}{\tau}\left((1+\gamma)(w^{\mathrm{D1}}_{\mathrm{FC}} x_i^{\mathrm{FC}} - w^{\mathrm{D1}}_{\mathrm{GPe}} x_i^{\mathrm{GPe}} + w^{\mathrm{D1}}_{S} S_i(t)) - w^{\mathrm{D1}}_{\mathrm{FS}} x^{\mathrm{FS}} + I_{\mathrm{D1}}\right) \quad (6)$$

$$(\mathbf{Wx} + \mathbf{I}(t))_{\mathrm{D2}_i} = \frac{1}{\tau}\left((1-\gamma)(w^{\mathrm{D2}}_{\mathrm{FC}} x_i^{\mathrm{FC}} - w^{\mathrm{D2}}_{\mathrm{GPe}} x_i^{\mathrm{GPe}} + w^{\mathrm{D2}}_{S} S_i(t)) - w^{\mathrm{D2}}_{\mathrm{FS}} x^{\mathrm{FS}} + I_{\mathrm{D2}}\right) \quad (7)$$

Table 1
Parameters of the simulations

N	6	τ	40 ms	τ_{STN}	5 ms	τ_{FS}	5 ms	τ_{FC}	80 ms
τ_{TH}	5 ms	τ_{TRN}	5 ms	γ	0.2	w^{D2}_{GPe}	1	w^{GPe}_{D2}	0.4
w^{D1}_{GPe}	1	w^{GPe}_{D1}	0.4	w^{FS}_{GPe}	0.05	w^{D1}_{FS}	0.5	w^{D2}_{FS}	0.5
w^{GPe}_{STN}	0.7	w^{STN}_{GPe}	0.45	w^{GPi}_{GPe}	0.08	w^{GPi}_{STN}	0.7	w^{GPi}_{D1}	0.4
w^{TH}_{TRN}	0.35	w^{TRN}_{TH}	0.35	w^{TH}_{FC}	0.6	w^{FC}_{TH}	0.6	w^{TRN}_{FC}	0.35
w^{TH}_{GPi}	0.18	w^{STN}_{FC}	0.58	w^{D1}_{FC}	0.1	w^{D2}_{FC}	0.1	w^{FS}_{FC}	0.01
I_{D1}	−0.1	I_{D2}	−0.1	I_{STN}	0.5	I_{GPi}	0.1	I_{GPe}	0.1

where $S(t)$ is the salience input vector, and where the negative constant inputs I_{D1} and I_{D2}, which keep the neurons silent when the inputs are not strong enough, model the up-state/down-state property of the MSNs.

The striatum also contains a small proportion of phenotypically diverse interneurons (Tepper & Bolam, 2004). We include here the fast spiking GABAergic interneurons (FS), that we model roughly as a single population exerting feedforward inhibition on the MSN (Tepper, Koós, & Wilson, 2004), and modulated by GPe feedback (Bevan et al., 1998)

$$(\mathbf{Wx}+\mathbf{I}(t))_{FS} = \frac{1}{\tau_{FS}}\sum_{j=1}^{N}\left(w^{FS}_{FC}x^{FC}_j - w^{FS}_{GPe}x^{GPe}_j + w^{FS}_{S}S_j(t)\right). \quad (8)$$

The subthalamic nucleus (STN) is the second input of the basal ganglia and also receives diffuse projections from the GPe, as explained above. Its glutamatergic neurons have an excitatory effect and project to the GPe and GPi. The resulting input of the STN neuron is given by

$$(\mathbf{Wx}+\mathbf{I}(t))_{STN_i} = \frac{1}{\tau_{STN}}\left(w^{STN}_{FC}x^{FC}_i - w^{STN}_{GPe}\sum_{j=1}^{N}x^{GPe}_j + I_{STN}\right) \quad (9)$$

where the constant positive input I_{STN} models the tonic activity of the STN.

The GPe is an inhibitory nucleus, it receives channel-to-channel afferents from the whole striatum (Wu, Richard, & Parent, 2000), and a diffuse excitation from the STN

$$(\mathbf{Wx}+\mathbf{I}(t))_{GPe_i} = \frac{1}{\tau}\left(-w^{GPe}_{D1}x^{D1}_i - w^{GPe}_{D2}x^{D2}_i + w^{GPe}_{STN}\sum_{j=1}^{N}x^{STN}_j + I_{GPe}\right) \quad (10)$$

where the constant positive input I_{GPe} models the tonic activity of the GPe.

The GPi and SNr are the inhibitory output nuclei of the BG, which keep their targets under inhibition unless a channel is selected. They receive channel-to-channel projections from the D1 striatum and diffuse projections from the STN and the GPe

$$(\mathbf{Wx}+\mathbf{I}(t))_{GPi_i} = \frac{1}{\tau}\left(-w^{GPi}_{D1}x^{D1}_i + w^{GPi}_{STN}\sum_{j=1}^{N}x^{STN}_i - w^{GPi}_{GPe}\sum_{j=1}^{N}x^{GPe}_j + I_{GPi}\right) \quad (11)$$

where the constant positive input I_{GPi} models the tonic activity of the GPi/SNr.

Finally, the thalamus (TH) forms an excitatory loop with the frontal cortex (FC), these two modules representing different thalamus nuclei and cortical areas, depending on the cortico-baso-thalamo-cortical loop considered. The thalamus is moreover under a global regulatory inhibition of the thalamic reticular nucleus (TRN, represented by a single population of neurons) and a channel-specific selective inhibition from the basal ganglia

$$(\mathbf{Wx}+\mathbf{I}(t))_{TH_i} = \frac{1}{\tau_{TH}}\left(w^{TH}_{FC}x^{FC}_i - w^{TH}_{TRN}x^{TRN} - w^{TH}_{GPi}x^{GPi}_i\right) \quad (12)$$

$$(\mathbf{Wx}+\mathbf{I}(t))_{FC_i} = \frac{1}{\tau_{FC}}\left(w^{FC}_{S}S_i + w^{FC}_{TH}x^{TH}_i\right) \quad (13)$$

$$(\mathbf{Wx}+\mathbf{I}(t))_{TRN} = \frac{1}{\tau_{TRN}}\left(\sum_i w^{TRN}_{FC}x^{FC}_i + w^{TRN}_{TH}x^{TH}_i\right). \quad (14)$$

This model keeps the basic off-center on-surround selecting structure, duplicated in the D1-STN-GPi/SNr and D2-STN-GPe subcircuits, of the GPR. However, the channel-specific feedback from the GPe to the Striatum helps in sharpening the selection by favoring the channel with the highest salience in D1 and D2. Moreover, the global GPe inhibition on the GPi/SNr synergetically interacts with the STN excitation in order to limit the amplitude of variation of the inhibition of the unselected channels. The inhibitory projections of the BG onto the thalamo-cortical excitatory loop limits the amplification of the unselected channels and thus favors a selective amplification of the winning channels. In such an architecture, the frontal cortex preserves the information from all channels but amplifies selectively the winning channel, in a sort of attention "spotlight" process, while the subcortical target circuits of the BG are under very selective inhibition, ensuring that motor commands do not interfere.

4. Disembodied model results

We first analyze the contraction of the contracting basal ganglia model (CBG) and its selection properties in simple disembodied tests before evaluating it as an ASM in a simulated robot.

Similarly to the simulations made by Gurney et al. (2001b), we used a 6-channel model. The parameters of the model were hand-tuned in order to obtain a selective system and respecting the local contraction constraints defined below, their values are summarized in Table 1. The simulation was programmed in C++, using the simple Euler approximation for integration, with a time step of 1 ms.

4.1. Contraction analysis of the model

According to the theory developed in Section 2.3, our model is contracting if the non-projected dynamics (which are linear) are contracting in a diagonal metric. To find this metric, we will use the three combinations presented in Section 2.4 that preserve diagonality.

Remark that each separated nucleus is trivially contracting in the identity metric because there is no lateral connection. The contracting rate of each nucleus is $\frac{1}{\tau}$, where τ is the common time constant of the N neurons of the nucleus. Thus, the metric $\mathbf{M}_{BG}$ of the basal ganglia is constituted of the blocks $\kappa_{GPe}\mathbf{I}$, $\kappa_{STN}\mathbf{I}$, $\kappa_{D1}\mathbf{I}$, $\kappa_{D2}\mathbf{I}$, $\kappa_{FS}1$ and $\kappa_{GPi}\mathbf{I}$. Similarly, the thalamic metric $\mathbf{M}_{TH}$ is constituted of

the blocks $\kappa_{\mathrm{FC}}\mathbf{I}$, $\kappa_{\mathrm{TH}}1$ and $\kappa_{TRN}\mathbf{I}$. The resulting metric for the whole system $\mathbf{M}_{\mathrm{CBG}}$ combines $\mathbf{M}_{BG}$ and $\mathbf{M}_{\mathrm{TH}}$ in the following way

$$\mathbf{M}_{\mathrm{CBG}} = \begin{pmatrix} \mathbf{M}_{BG} & 0 \\ 0 & \alpha\mathbf{M}_{\mathrm{TH}} \end{pmatrix}.$$

4.1.1. Analysis of the basal ganglia

- $\kappa_{\mathrm{GPe}} = 1$.
  We can set κ_{GPe} to any value as there is no combination at this stage. The current contracting rate is $\frac{1}{\tau}$.
- $\kappa_{\mathrm{STN}} = w^{\mathrm{GPe}}_{\mathrm{STN}}/w^{\mathrm{STN}}_{\mathrm{GPe}}$.
  We use negative feedback. The contracting rate remains unchanged
- $\begin{cases} \kappa_{\mathrm{D1}} = w^{\mathrm{GPe}}_{\mathrm{D1}}/((1+\gamma)w^{\mathrm{D1}}_{\mathrm{GPe}}) \\ \kappa_{\mathrm{D2}} = w^{\mathrm{GPe}}_{\mathrm{D2}}/((1-\gamma)w^{\mathrm{D2}}_{\mathrm{GPe}}). \end{cases}$
  We use small gains to show that the system constituted by the STN, GPe, striatum D1 and D2 is contracting when

$$\left((1+\gamma)w^{\mathrm{GPe}}_{\mathrm{D1}}w^{\mathrm{D1}}_{\mathrm{GPe}}\right)^2 + \left((1-\gamma)w^{\mathrm{GPe}}_{\mathrm{D2}}w^{\mathrm{D2}}_{\mathrm{GPe}}\right)^2 < 1 \tag{15}$$

  with a contracting rate

$$\frac{1}{\tau}\left(1 - \sqrt{((1+\gamma)w^{\mathrm{GPe}}_{\mathrm{D1}}w^{\mathrm{D1}}_{\mathrm{GPe}})^2 + ((1-\gamma)w^{\mathrm{GPe}}_{\mathrm{D2}}w^{\mathrm{D2}}_{\mathrm{GPe}})^2}\right).$$

- $\kappa_{\mathrm{FS}} = w^{\mathrm{D1}}_{\mathrm{FS}}/w^{\mathrm{FS}}_{\mathrm{GPe}}$.
  Again by use of small gains.
- $\kappa_{\mathrm{GPi}} = 1/(\tau\sigma(\mathbf{G}))^2$

where $\sigma(\mathbf{G})$ is the largest singular value of the matrix of projections on GPi and τ is the slowest time constant of neurons in the basal ganglia. This constant is set by using hierarchical combination.

Thus we can guarantee the contraction of the basal ganglia as soon as condition (15) is satisfied.

4.1.2. Analysis of the thalamus

- $\kappa_{\mathrm{TH}} = 1$.
  We can set κ_{TH} to any value as there is no combination at this stage. The current contracting rate is $\frac{1}{\tau_{\mathrm{TH}}}$.
- $\kappa_{\mathrm{GPe}} = w^{\mathrm{TH}}_{TRN}/w^{TRN}_{\mathrm{TH}}$.
  We use negative feedback. The contracting rate remains unchanged
- $\kappa_{\mathrm{FC}} = \sqrt{{w^{\mathrm{TH}}_{\mathrm{FC}}}^2 + N{w^{TRN}_{\mathrm{FC}}}^2}/w^{\mathrm{FC}}_{\mathrm{TH}}$.
  We use small gains to show that the thalamo-cortical module is contracting when

$$w^{\mathrm{FC}}_{\mathrm{TH}}\left(w^{\mathrm{TH}}_{\mathrm{FC}} + \sqrt{{w^{\mathrm{TH}}_{\mathrm{FC}}}^2 + N{w^{TRN}_{\mathrm{FC}}}^2}\right) < 1. \tag{16}$$

  Remark that this condition depends on N. This would not have been the case if we had modeled the *TRN* by N channels instead of 1.

Thus we can guarantee the contraction of the thalamus as soon as condition (16) is satisfied.

It remains to examine the large loop between the thalamus and the basal ganglia involving projections of the GPi and the FC. Again, we use small gains to set α.

$$\alpha = \sqrt{\frac{\tau_{FCtx}\kappa_{\mathrm{GPi}}\,({w^{\mathrm{STN}}_{\mathrm{FC}}}^2 + {w^{\mathrm{D1}}_{\mathrm{FC}}}^2 + {w^{\mathrm{D2}}_{\mathrm{FC}}}^2 + n{w^{\mathrm{FS}}_{\mathrm{FC}}}^2)}{\tau_{\mathrm{TH}}\kappa_{\mathrm{FC}}\,{w^{\mathrm{TH}}_{\mathrm{GPi}}}^2}}.$$

Proposition 2. *Let* $\mathbf{M}_{\mathrm{CBG}} = \mathbf{\Theta}^{\mathrm{T}}_{\mathrm{CBG}}\mathbf{\Theta}_{\mathrm{CBG}}$ *be the diagonal metric defined above. By Theorem 2, if the generalized Jacobian* $\mathbf{\Theta}_{\mathrm{CBG}}\mathbf{W}\mathbf{\Theta}^{-1}_{\mathrm{CBG}}$ *is negative definite, the dynamical system* $\dot{\mathbf{x}} = \mathbf{\Pi}_{\mathbb{H}_n}(\mathbf{x}, \mathbf{W}\mathbf{x} + \mathbf{I}(t))$ *describing the cortico-baso-thalamo-cortical loop model is contracting with a rate* $|\lambda_{\max}|$*, where* $\lambda_{\max}$ *is the largest eigenvalue of* $\mathbf{\Theta}_{\mathrm{CBG}}\mathbf{W}\mathbf{\Theta}^{-1}_{\mathrm{CBG}}$.

Table 2
Value of the constants defining the metric M_{CBG} for the set of parameters of our simulation

κ_{GPe}	κ_{STN}	κ_{D1}	κ_{D2}	κ_{FS}	κ_{GPi}	κ_{TH}	κ_{TRN}	κ_{FCtx}	α
1	0.441	0.577	0.707	1	0.104	1	1	5.282	0.253

At this stage, we have provided an algebraic definition of the metric $\mathbf{M}_{\mathrm{CBG}}$. Unfortunately, the complexity of the induced generalized Jacobian prevents us from giving a global algebraic condition on the projection weights for the generalized Jacobian to be negative definite. This is not of major incidence as we can compute numerically, for any instance of the weights, the eigenvalues of the symmetric part of the generalized Jacobian and check that they are all negative.

Table 2 gives the numerical value of the constants defining the metric M_{CBG} for the set of parameters of our simulation (see Table 1). Using the free software *Octave*, we compute in that case the eigenvalues of the generalized Jacobian and obtain that our model is contracting with contracting rate of 2.20.

Notice that computing the maximum real part of the eigenvalue of the non-projected dynamics (which are linear) gives an upper bound of the contracting rate. For the set of parameters of our simulation, this upper bound is 2.59. It is remarkable that being forced to use diagonal metrics in our proof (which discards a huge set of metrics) has not decreased much the contracting rate.

4.2. Basic selection test

We first reproduced the selection test of Gurney et al. (2001b) with our model and with the GPR model version presented in Prescott et al. (2006). In this test, a specific sequence of five different salience vectors (represented by the dashed lines in Fig. 3) is submitted to a 6-channel version of the BG model, in order to show the basic selection properties of the system. Here, we submitted each vector to the system during 2 s before switching to the next one in the sequence.

During the CBG simulation (Fig. 3, top row), with the first vector of null saliences, the system stabilizes in a state where all channels are equally inhibited ($x^{\mathrm{GPi}}_i = 0.095$). Then, the first channel receives a 0.4 input salience which results in a clear disinhibition of this channel ($x^{\mathrm{GPi}}_1 = 0.014$) and increased inhibition of the others. When the second channel salience is set to 0.6, it becomes perfectly selected ($x^{\mathrm{GPi}}_2 = 0$) while the first one is rapidly inhibited to a level identical to the one of the four last channels. During the fourth step, the salience of the first channel is increased to 0.6, channels 1 and 2 are therefore simultaneously selected ($x^{\mathrm{GPi}}_1 = x^{\mathrm{GPi}}_2 = 0.03$). Finally, during the last step of the test, channel 1 has its salience reduced to 0.4, and it is then rapidly inhibited, while channel 2 returns to perfect selection ($x^{\mathrm{GPi}}_2 = 0$). The CBG thus passes this test in a satisfactory manner: the channels with the highest saliences are always selected while the others are inhibited.

The GPR simulation (Fig. 3, bottom row) is qualitatively quite similar, excepted during the fourth step of the sequence (emphasized with an asterisk): while the salience of channel 1 increases from 0.4 up to 0.6 (the same salience as that of channel 2), channel 2 remains selected and channel 1 is fully inhibited (its level of inhibition is higher than the inhibition at rest). The inputs in channels 1 and 2 being exactly the same, this difference in their selection state is clearly caused by the initial conditions of the system (i.e. the fact that channel 2 was selected before). This example of a dependence on the initial conditions clearly shows that the GPR model is not contracting.

Indeed, as we have seen in Section 2.3, a rate-coding neural network is contracting only if its non-projected dynamics are contracting in a diagonal metric. But a linear system is stable if and

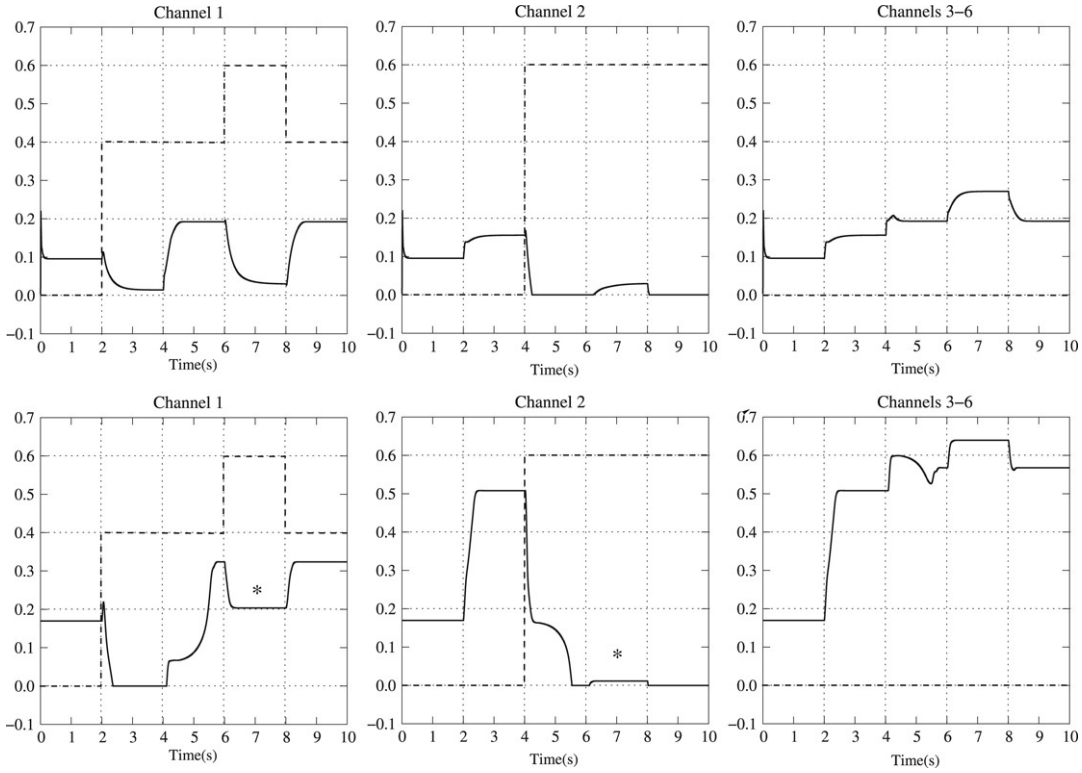

Fig. 3. Variation of the GPi/SNr inhibitory output during the Gurney et al. (2001b) test applied to (top) the CBG and (bottom) the GPR. Dashed lines represent the input salience of the channel and solid lines represent the output of the channel. Note that during the fourth step ($6\ s < t < 8\ s$), channels 1 and 2 are selected by the CBG, while the GPR selects channel 2 only (asterisk).

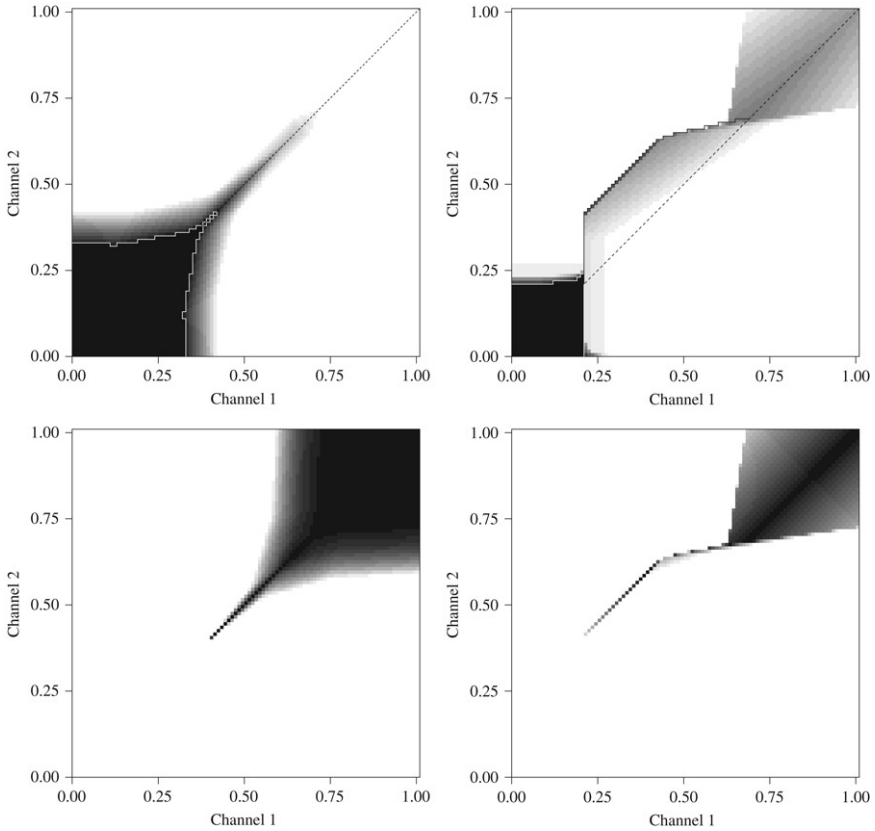

Fig. 4. Efficiency (top) and distortion (bottom) in the winning channel for a systematic salience-space search for the CBG (left) and the GPR (right). Top: black to white gradient represents increasing efficiency (from 0 to 1); bottom: black to white gradient represents decreasing distortion (from 1 to 0), maximal distortion corresponding to simultaneous selection of both channels is thus in black. White line: limit beyond which no selection occurs; dashed black line: diagonal representing equal saliences. For the GPR efficiency (top right), note the hysteresis area between the dashed and the full black lines. See the text for further explanations.

only if all its eigenvalues have a negative real part. Computing the eigenvalues of the linear part of the GPR reveals that $N-1$ of them have a positive real part (namely 10.387). We can thus conclude that the GPR is not contracting.

4.3. Systematic salience search test

This first result is however not surprising, as revealed by the systematic salience search experiment performed in Prescott et al. (2006), and that we also reproduced with both the GPR and the CBG. In this experiment, the first two channels of the ASM are put in competition in the following manner: the first channel salience is increased from 0 to 1 in steps of 0.01, and for each of these steps, the salience of the second channel is also gradually increased from 0 to 1 in steps of 0.01. The system is run to convergence between all step increases. The internal state of the model is not reset between each channel 2 salience increase, but only for channel 1 steps. This means that the test evaluates the selection response of the system with one channel salience fixed while the other one gradually increases.

In order to evaluate the response of the ASM to this experiment, four numerical values are computed. First, the efficiencies of the selection of channels 1 and 2, equivalent to the percentage of disinhibition, are computed as follows:

$$e_i = [1 - y_i^{\mathrm{GPi}}/y_{\mathrm{Rest}}^{\mathrm{GPi}}]_+ \tag{17}$$

with i the index of the channel, y_i^{GPi} the output of the ith GPi neuron and $y_{\mathrm{Rest}}^{\mathrm{GPi}}$ the output inhibition of all channels when all saliences are null. The absolute efficiency of the selection is defined as the efficiency of the winning channel:

$$e_w = \max_i e_i. \tag{18}$$

Finally, the distortion of the selection, which is null when only the winning channel is disinhibited and increasing with the disinhibition of its competitors, is defined by:

$$d_w = 2\frac{\sum_i e_i - e_w}{\sum_i e_i}. \tag{19}$$

The results of the experiment are summarized by the e_w and d_w graphs (Fig. 4), where the value of each of these variables is represented with regard to the corresponding channel 1 (abscissa) and channel 2 (ordinate) saliences. First observe that the GPR results we obtain with 6 channels are very similar to those presented in Prescott et al. (2006) for a 5-channel GPR. Concerning e_w (top row), whereas, for the CBG, the selection switches from channel 1 to channel 2 as soon as the salience of channel 2 is larger than the salience of channel 1 (when it crosses the diagonal in dashed black), for the GPR, this switch is delayed until much higher values are reached (when it crosses the black line). As previously noted, this hysteresis effect is a direct consequence of the non-contraction of the GPR.

Note that when high saliences are in competition, the GPR tends to partially select both channels ($e_w < 1$ and $d_w > 0$), while the CBG fully disinhibits both channels ($e_w = 1$ and d_w close to 1). Which behavior is preferable for an ASM is not decided.

Is the GPR's strong dependence on initial conditions a good feature for an ASM? Prescott et al. (2006) argue that it allows behavioral persistence, and that in their experiment, the robot takes advantage of it to avoid dithering between actions. We do not claim that there is a definitive answer to the question. Nevertheless, in the next section, we describe the evaluation of the CBG in a minimal survival task in which the robot also avoids dithering, despite its contracting ASM. This shows that this dependence on initial conditions is not necessary from the point of view of dithering avoidance.

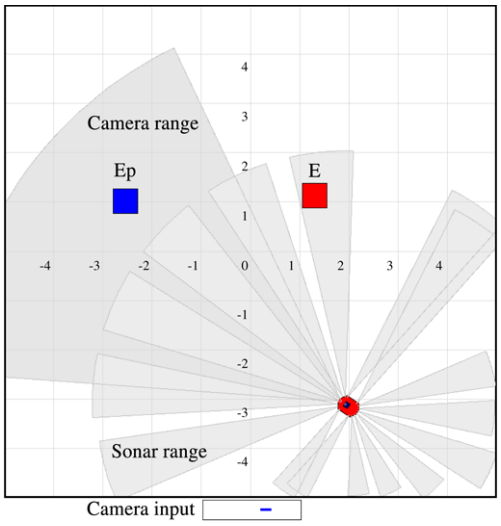

Fig. 5. Experimental set-up. Blue square: Potential Energy resource; red square: Energy resource. The light gray surfaces represent the field of view of the sonars, and the darker one the field of view of the camera. The corresponding camera image is represented at the bottom.

5. Minimal survival task

5.1. Materials and methods

The suitability of the model for action selection in an autonomous robot has been tested in simulation with the same minimal survival task previously used to evaluate the GPR model (Girard et al., 2003). In order to emphasize its properties, and in particular those resulting from the selective feedback loop, its performance was compared to a simple if-then-else decision rule (ITE, fully described in Appendix A).

In such a task, the robot has to go back and forth between locations containing two different kind of resources, in order to keep its energy level above 0. The robot has two internal variables, namely *Energy* and *Potential Energy*, taking values between 0 and 1, and an artificial metabolism, which couples them as follows:

- The Energy (E) is continuously decreasing, with a constant consumption rate (0.01 Energy unit per second). When it reaches 0, the robot has run out of energy and the ongoing trial is interrupted. To prevent this, the robot has to regularly acquire Energy by activating the *ReloadOnE* action on an Energy resource. Note that ReloadOnE only *transforms* Potential Energy into Energy (0.2 units of E_p are transformed into 0.2 units of E each second), thus Potential Energy has to be also reloaded.
- The Potential Energy (E_p) is a sort of Energy storage, it can be acquired by activating the *ReloadOnEp* action on a Potential Energy resource, and is consumed in the transformation process only.

In this version of the task, the experiments are run in simulation using the Player/Stage robot interface and robot simulator (Gerkey, Vaughan, & Howard, 2003). The simulated robot is a 40 × 50 cm wheeled robot with differential steering, similar to the Activ-Media Pioneer 2DX (Fig. 5), equipped with a ring of 16 sonars and a camera. The sonar sensors have a maximum range of 5 m and a view angle of 15°, the camera has a resolution of 200 × 40 pixels and a view angle of 60° and uses a color-blob-finding vision device to track the position of red and blue objects. The experiment takes place in a 10 × 10 m arena, containing one Energy and one Potential Energy resource (Fig. 5). These resources are represented by colored 50 × 50 cm objects (respectively red and blue), and do not constitute obstacles (as if they were suspended above the

arena). They are randomly positioned in the arena for each trial, with the constraint that their center is at least 1 m away from the walls.

The robot has to select from among seven possible actions:

- ReloadOnE (ROE) and ReloadOnEp (ROE$_p$) affect the robot's survival as previously described. These actions are effective if the robot is facing the corresponding resource and is close enough (45° of the camera field of view is occupied by the resource).
- Wander (W) activates random accelerations, decelerations and turning movements.
- Rest (R) stops the robot, which is a disadvantage as the robot has to continuously explore the arena to find resources, but Rest also halves the rate of Energy consumption (0.005 unit per s), which promotes long survival. Consequently, it should be activated when there is no risk (i.e. when both internal variables reach high levels) in order to minimize the Potential Energy extracted from the environment to survive.
- AvoidObstacle (AO) uses data from the 6 front sonars and the 2 central rear sonars in order to avoid collisions with walls.
- ApproachE (AE) and ApproachEp (AE_p) use the color-blob-finder in order to orient and displace the robot towards the corresponding resource if it is visible.

The action selection mechanisms base their decisions on the following variables:

- E, E_p,$(1-E)$ and $(1-E_p)$, which provide the amount (or lack of) Energy and Potential Energy,
- *seeEBlob* and *seeEpBlob*, which are set to 1 if a red (resp. blue) object is in the camera input, and to 0 otherwise,
- *onEBlob* and *onEpBlob*, which are set to 1 if a red (resp. blue) object is larger than 150 pixels (i.e. close enough to allow the use of the corresponding resource), and to 0 otherwise,
- *SFR* and *SFL* are the values of the front-right and front-left sonar sensors, measured in meters, taking values between 0 and 5.

For the CBG, the detailed salience computation using these variables is given in Appendix B.

The action selection mechanisms receive new sensory data every 100 ms, and must then provide an action selection for the next 100 ms. Concerning the ITE, it is simply done by executing the decision rule once with the latest data. Concerning the CBG, the selection is made using the output inhibition resulting from the computation of 100 simulation steps of 1ms, using the latest sensory data. A given action is then considered selected if the inhibition of the corresponding channel is below the inhibition at rest $y^{\mathrm{GPi}}_{\mathrm{Rest}}$ (as defined previously). In the case of multiple channel disinhibition, the following action combination rules have been defined:

- Rest is effective if and only if it is the only disinhibited action,
- ReloadOnE and ReloadOnEp are effective if and only if the robot does not move,
- The other movement-generating actions can be co-activated. In that case, the efficiency of selection (as defined by Eq. (17)) is used to weight the contributions of each action to the final motor command.

The comparison between the CBG and the ITE is made according to the following protocol: 20 random resource positions are drawn and, for each model, 20 trials are run using the same set of positions. The robot begins the experiment with a full battery ($E = 1$) and no Potential Energy storage ($E_p = 0$), this allows a maximal survival duration of 1 min 40 s if no reloading action occurs. Unless the robot runs out of energy ($E = 0$), the trial is stopped after 15 min.

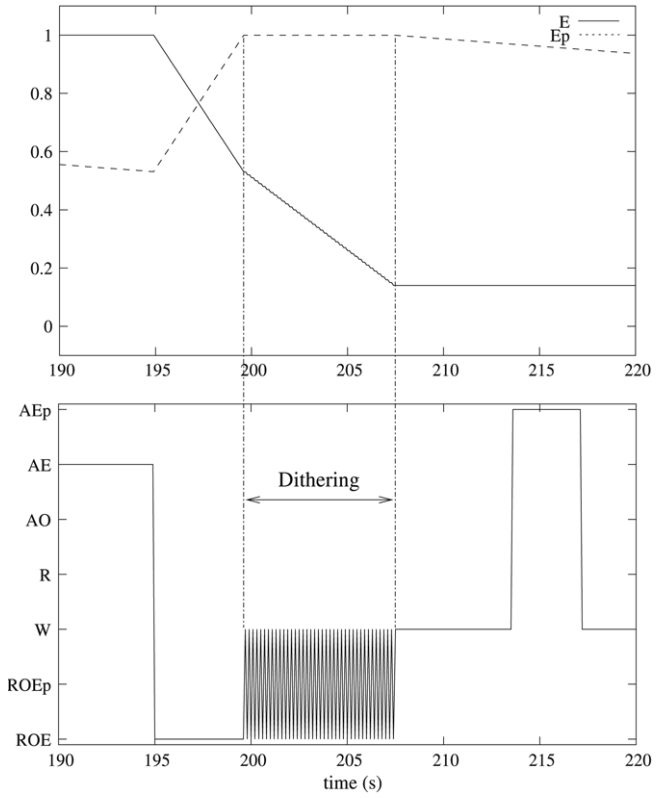

Fig. 6. Typical dithering of the ITE between the ReloadOnEnergy and Wander actions. Top: levels of Energy (dashed line) and Potential Energy (full line); bottom: selected action. Note how during the dithering period, more than 0.3 units of E_p are wasted in about 7 s, while they should have allowed 30 s of survival.

5.2. Results

The first result is that the CBG and the ITE algorithm have similar survival performance. They are both able to survive the trial in a majority of cases, but can be subject to premature Energy shortage. This is expected, because their ability to find resources is limited by the camera range and field of view, as well as by the random exploration action. The average survival duration is 687 s ($\sigma = 244$) for the CBG and 737 s ($\sigma = 218$) for the ITE, and the two-tailed Kolmogorov–Smirnov test confirms that the two sets of survival durations are not drawn from significantly different distributions ($D_{KS} = 0.2, p = 0.771$). From an action selection point of view, the comparison of the two mechanisms is thus fair: despite the fact that they were tuned independently, they both achieve similar survival performance.

Nevertheless, *a clear behavioral difference between the two mechanisms was observed*, which has significant repercussions on their ability to store Potential Energy and on the Potential Energy extracted from the environment. Indeed, while the CBG may use its feedback loops in order to persist in action execution, the ITE was deliberately deprived of any memory. This was done in order to investigate the effects of this persistence property. The ITE exhibits behavioral dithering in a critical and frequent situation: when the robot fully reloads its Energy, it activates the Wander action, but after 100 ms of Wander execution, some Energy has been consumed and the robot has not moved much. In most cases, it is still on the Energy resource, and if it still has spare E_p, ReloadOnE is activated again. This repeats until there is no E_p left or until, in a sequence of small movements, the robot has left the resource (see Fig. 6). This dithering generates a strong energy dissipation: 100 ms of Wander consumes 0.001 units of Energy, and during the following 100 ms, ReloadOnEnergy consumes 0.02 units of E_p while E, being bounded by 1, increases by 0.001 only.

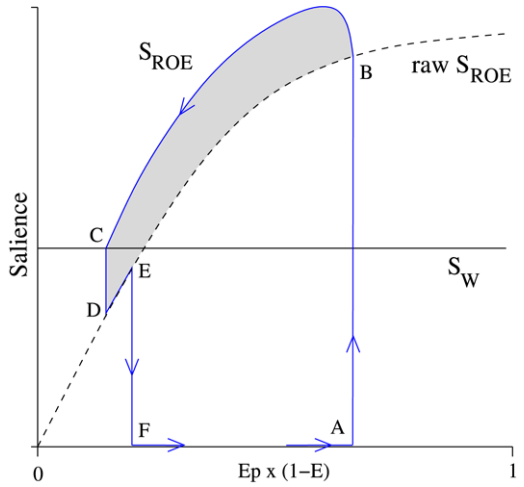

Fig. 7. Hysteresis in the variation of the salience of ReloadOnEnergy for the CBG. Black dashed line: variation of S_{ROE} with regard to $(E_p \times (1 - E))$, with $onEBlob = 1$ and without the persistence term (raw S_{ROE}); blue line: variation of S_{ROE}; shaded area: S_{ROE} increase resulting from the frontal cortex feedback; black line: salience of Wander (S_W). Explanations are given in the text.

On the contrary, in the same situation, the CBG takes advantage of a hysteresis effect caused by the positive feedback from the frontal cortex to the basal ganglia to avoid dithering.

Indeed, the salience of ROE is defined by: $S_{\text{ROE}} = 950 \times f(4 \times onEBlob \times E_p \times (1 - E)) + 0.6 \times x^{\text{FC}}_{\text{ROE}}$ (where f is a sigmoid transfer function, see Appendix B). Consequently, when the robot has a lack of Energy and reaches an Energy resource, *onEBlob* jumps from 0 to 1 and S_{ROE} also jumps from 0 (Fig. 7, point A) to a level depending on the current E and E_p internal states (Fig. 7, point B) situated on the raw S_{ROE} curve (Fig. 7, dashed line). In the case depicted in Fig. 7, S_{ROE} is then much higher than S_W, and ROE is thus selected. As a consequence, the corresponding thalamo-cortical channel is disinhibited, leading to an amplification of the salience, fed back to the basal ganglia thanks to the cortical output $x^{\text{FC}}_{\text{ROE}}$ (this bonus is represented by the shaded area over the raw S_{ROE} curve on Fig. 7).

While the robot reloads, S_{ROE} decreases with $(E_p \times (1 - E))$, but because of the $x^{\text{FC}}_{\text{ROE}}$ salience bonus, it follows the blue trajectory down to point C, where Wander is selected again. The deselection of ROE shuts off the $x^{\text{FC}}_{\text{ROE}}$ signal, causing an immediate decrease to point D. As soon as the robot activates Wander, Energy is consumed and S_{ROE} increases again, along the raw S_{ROE} curve. However, at point D, $S_{\text{ROE}} < S_W$, and as long as the robot manages to leave the resource before S_{ROE} exceeds S_W (points E and F, when the *OnEBlob* variable jumps from 1 to 0), no dithering occurs.

This observation is not trivial, as it has a direct consequence on the global E_p storage of the ITE: both CBG and ITE keep high levels of E_p (between 0.9 and 1) more than 50% of the time (Fig. 8, right), but for the rest of the time, the ITE level is very low (0–0.1) much more often (almost 20% of the time) than the CBG. Moreover, the CBG activates the Rest action often enough to extract, on average, less Potential Energy from the environment (0.93×10^{-2}Ep s^{-1}, $\sigma = 0.30 \times 10^{-3}$) than the basic rate ($1 \times 10^{-2}$Ep s^{-1}). On the contrary, the dissipation of energy caused by the dithering of the ITE generates a much higher Potential Energy extraction rate (1.17×10^{-2}Ep s^{-1}, $\sigma = 1.17 \times 10^{-3}$). The two-tailed Kolmogorov–Smirnov test reveals that the E_p consumption rates measured for the CBG and the ITE (Fig. 9) are drawn from different distributions ($D_{KS} = 0.95$, $p < 0.001$). The ITE dithering thus generates so much dissipation that it has to extract extra Potential Energy from the environment, despite its use of the Sleep action to lower its consumption, while the CBG exploits as much as possible this possibility to limit Potential Energy extraction.

6. Discussion

We proposed a new action selection mechanism for an autonomous robot, using a multi-disciplinary approach combining computational neuroscience and dynamic system theory. This study proved fruitful in the three considered domains:

- We proposed an extension of the contraction theory to locally projected dynamical systems, which was necessary to study the stability of rate-coding neural networks.

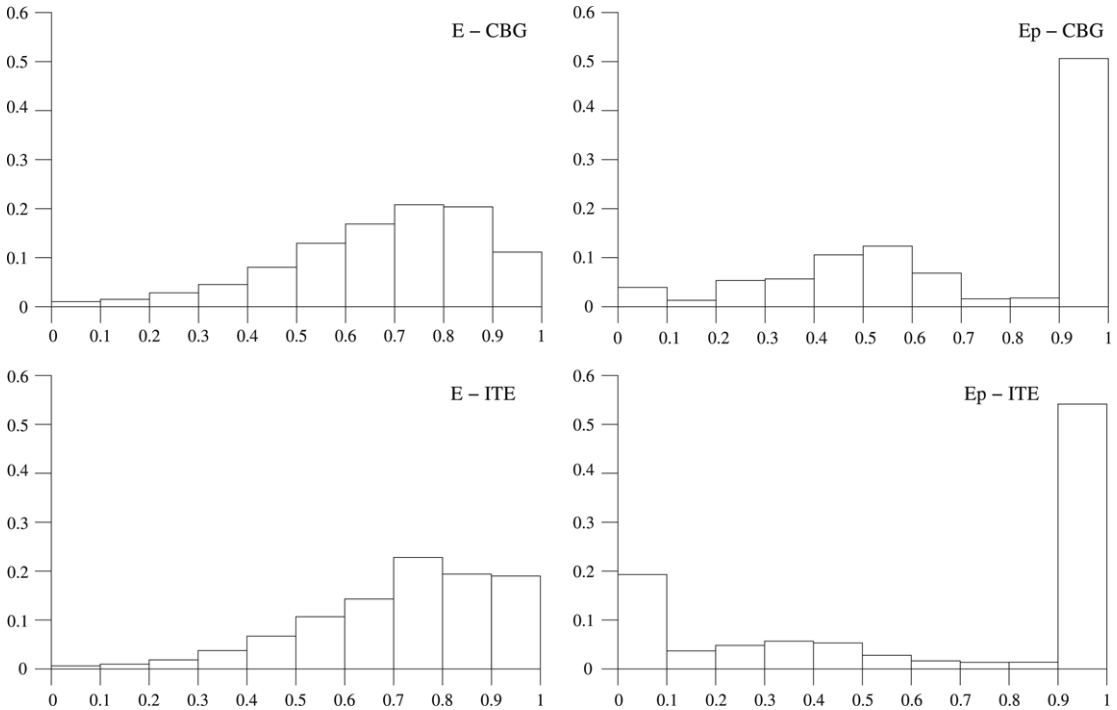

Fig. 8. Histograms of Energy (left) and Potential Energy (right) for the CBG (top) and the ITE (bottom), cumulated over all trials.

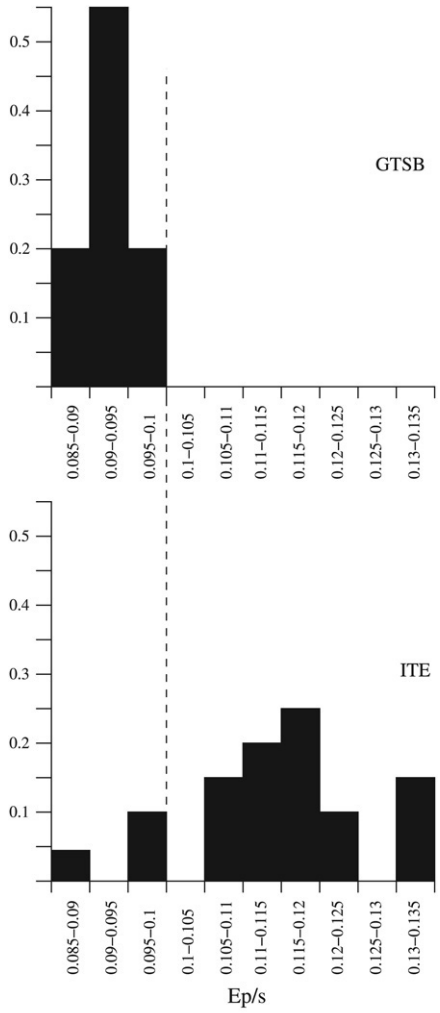

Fig. 9. Potential Energy consumption rate. These histograms represent the average E_p consumption rate computed for each trial. Top: BG model; bottom: ITE; the dashed line shows the Energy consumption rate of all actions except Rest (0.001 E/s).

- As a consequence, we proposed a modified rate-coding artificial neuron model.
- Using these results, we designed a stable model of the cortico-baso-thalamo-cortical loops (CBG) using previously neglected anatomical data.
- After having tested this model offline, we integrated it in a simulated robot confronted to a standard survival task to assess its efficiency as an action selection mechanism.

6.1. Dynamic systems

In this paper, we have investigated the stability properties of locally projected dynamical systems (lPDS) using nonlinear contraction theory. In particular, we have given a sufficient condition for a general non-autonomous (i.e. with time-varying inputs) lPDS to be *globally exponentially stable*. By contrast, Zhang and Nagurney (1995) only studied the stability of a fixed equilibrium point in autonomous lPDS. Thus, the novelty of our theoretical result should be noticed.

Locally projected dynamical systems have attracted great interest since they were introduced in 1993 by Dupuis and Nagurney. Indeed, this theory is central to the study of oligopolistic markets, traffic networks, commodity production, etc (Dupuis & Nagurney, 1993). As we demonstrated in this article, this theory has also proved to be a valuable tool for establishing rigorous stability properties of neural networks. In this respect, further development of the theory as well as its application to numerous problems in theoretical neuroscience may represent exciting subjects of research.

6.2. Neuroscience

The CBG shares a number of similarities with the previously proposed GPR model (Gurney et al., 2001b), as its selection ability relies on two off-center on-surround subcircuits. However, it includes neglected connections from the GPe to the Striatum, which provide additional selectivity. It also considers the possible role of global projections of the GPe to the STN, GPi and SNr as a regulation of the activity in the whole basal ganglia.

We omitted two types of documented connections in the current CBG model. First, the STN projects not only to the GPe, GPi and SNr but also to the striatum (Parent et al., 2000). Intriguingly, the population of STN neurons projecting to the striatum does not project to the other targets, while the other STN neurons project to at least two of the other target nuclei (GPe, GPi or SNr). We could not decipher the role of this striatum-projecting population and did not include it in the current model. Its unique targeting specificity suggests it could be functionally distinct from the other STN neurons. To our knowledge, no modeling study has yet proposed a functional interpretation of this connection, a question that should be explored in future works. The other missing connections concern the fact that lateral inhibitions exist in GPe and SNr (Deniau, Kitai, Donoghue, & Grofova, 1982; Juraska, Wilson, & Groves, 1977; Park, Falls, & Kitai, 1982). These additional projections were added to a version of the GPR (Gurney, Humphries, Wood, Prescott, & Redgrave, 2004) and seemed to enhance its selectivity. We might add these connections and proceed to a similar test with the CBG.

The GPe to striatum connections have the previously evoked functional advantage of enhancing the quality of the selection, by silencing the unselected striatal neurons. Interestingly, the striatum is known for being a relatively silent nucleus (DeLong et al., 1984), a property supposed to be induced by the specific up/down state behavior of the striatal neurons. When using simple neuron models, like leaky integrators, it is usually difficult to reproduce this with a threshold in the transfer function only: when many channels have a strong salience input, all the corresponding striatal neurons tend to be activated. Our model suggests that in such a case, the GPe-striatum projections may contribute to silencing the striatum.

The proposed model includes the modulatory role of the dopamine (DA) in the BG selection process only, which corresponds to the tonic level of dopaminergic input from the ventral tegmental area and the substantia nigra pars compacta (VTA and SNc). The effects of the variation of this tonic DA level on the selection abilities of the BG has been examined in detail for the GPR (Gurney et al., 2001b), and compared with the symptoms of Parkinson's disease.

The role of the phasic dopamine activity in reinforcement learning, through the adaptation of the cortico-striatal synapses, is beyond the scope of our study. Nevertheless, such an extension of the CBG could allow the online adaptation of the saliences, which are here hand-tuned. The existing models of reinforcement learning in the BG are based on the temporal difference (TD) learning algorithm (Houk, Adams, & Barto, 1995; Joel, Niv, & Ruppin, 2002). These TD models are composed of two cooperating circuits: a *Critic* dedicated to learning to predict future reward given the current state, and an *Actor*, using the Critic's predictions to choose the most appropriate action. Our model can then be considered as an Actor circuit, more anatomically detailed than those usually used (simple winner-takes-all, without persistence properties). The first attempts at using detailed Actor models

in TD architectures for tasks requiring a single motivation have been conducted (Frank, Santamaria, O'Reilly, & Willcutt, 2007; Khamassi, Girard, Berthoz, & Guillot, 2004; Khamassi, Lachèze, Girard, Berthoz, & Guillot, 2005). Note however that the use of the current TD-learning models would not necessarily be straightforward in our case: we had to use relatively complex salience computations (see Appendix B), in order to solve our relatively simple task. This is caused by its multi-motivational nature, quite common in action selection problems, but which has been given only little attention in RL-related works (Dayan, 2001; Konidaris & Barto, 2006).

6.3. Autonomous robotics

While early action selection mechanisms were based on a purely engineering approach (Pirjanian, 1999), progress in the understanding of the physiology of the brain regions involved in action selection now allows the investigation of biomimetic action selection mechanisms. Indeed, basal ganglia models – variations of the GPR – and reticular formation models have already been used as action selection mechanisms for autonomous robots (Girard et al., 2003, 2005; Humphries, Gurney, & Prescott, 2005; Montes-Gonzalez et al., 2000; Prescott et al., 2006).

We showed here that the CBG may exploit its cortical feedback to exhibit behavioral persistence and thus dithering avoidance, one of the fundamental properties of efficient ASMs (Tyrrell, 1993). In our experiment, this promotes energy storage and reduces energy consumption. These properties, which clearly provide a survival advantage, were also highlighted for the GPR when tested in a similar experiment (Girard et al., 2003). Thus, comparing the GPR and the CBG in exactly the same task could reveal some subtle differences which were not identified yet. Moreover, in the current version of the CBG, these cortico-striatal feedback connections are strictly channel to channel, the possible sequence generation effects that could result from cross channel connections probably deserves additional attention.

The contraction property of the CBG also provides a fundamental advantage for an autonomous robot. It provides a theoretical certainty regarding its stability of operation, whatever the sequences of input might be. For an autonomous agent confronted with a uncontrolled environment, where all possible sequences of inputs may happen, it seems to be essential. Of course, contraction analysis does not say anything about the pertinence of the resulting stable behavior, hence leading the necessity of verifying the CBG selection properties. However, the fact that stability issues have already been evoked for previous GPR versions (Girard et al., 2005; Prescott et al., 2006) confirms that such a rigorous proof is useful.

Acknowledgment

B.G. and N.T. acknowledge the partial support of the European Community Neurobotics project, grant FP6-IST-001917.

Appendix A. If-Then-Else decision rule

The If-Then-Else decision tree is the following:

if $E_p < 1$ and $onEpBlob = true$ **then**
  ReloadOnEp
else if $E < 1$ and $E_p > 0$ and $onEBlob = true$ **then**
  ReloadOnE
else if $E < 0.8$ and $E_p > 0$ and $seeEBlob = true$ **then**
  ApproachE
else if $E_p < 0.8$ and $seeEpBlob = true$ **then**
  ApproachEp
else if $E > 0.7$ and $E_p > 0.7$ **then**
  Rest
else if $SFL < 1$ or $SFR < 1$ or ($SFL < 1.5$ and $SFR < 1.5$) **then**
  AvoidObstacle
else
  Wander
end if

Appendix B. Robot CBG saliences

Using the sigmoid transfer function

$$f(x) = \frac{2}{1 + e^{-4x}} - 1$$

the saliences of each action (including the frontal cortex feedback) are:

$$S_{\mathrm{ROE}} = 950 \times f(4 \times onEBlob \times E_p \times (1 - E)) + 0.6 \times x^{\mathrm{FC}}_{\mathrm{ROE}}$$
$$S_{\mathrm{ROE}_p} = 750 \times f(4 \times \mathrm{onEpBlob} \times (1 - E_p)) + 0.2 \times x^{\mathrm{FC}}_{\mathrm{ROE}_p}$$
$$S_W = 380$$
$$S_{SI} = 550 \times f(2 \times \max(Ep \times E - 0.5, 0))$$
$$S_{AO} = 950 \times f(2 \times (\max(1.5 - SFL, 0) + \max(1.5 - \mathrm{SFR}, 0))) + 0.2 \times x^{\mathrm{FC}}_{AO}$$
$$S_{AE} = 750 \times f(\mathrm{seeEBlob} \times E_p \times (1 - E) \times (1 - onEBlob)) + 0.2 \times x^{\mathrm{FC}}_{AE}$$
$$S_{AE_p} = 750 \times f(seeEpBlob \times (1 - E_p) \times (1 - \mathrm{onEpBlob})) + 0.2 \times x^{\mathrm{FC}}_{AE_p}.$$

References

Alexander, G. E., Crutcher, M. D., & DeLong, M. R. (1990). Basal ganglia-thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Progress in Brain Research*, *85*, 119–146.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*, 357–381.

Bevan, M., Booth, P., Eaton, S., & Bolam, J. (1998). Selective innervation of neostriatal interneurons by a subclass of neurons in the globus pallidus of rats. *Journal of Neuroscience*, *18*(22), 9438–9452.

Chevalier, G., & Deniau, M. (1990). Disinhibition as a basic process of striatal functions. *Trends in Neurosciences*, *13*, 277–280.

Dayan, P. (2001). Motivated reinforcement learning. In T. Leen, T. Dietterich, & V. Tresp (Eds.), *Neural information processing systems*: *Vol. 13*. Cambridge, MA: The MIT Press.

Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

DeLong, M., Georgopoulos, A., Crutcher, M., Mitchell, S., Richardson, R., & Alexander, G. (1984). Functional organization of the basal ganglia: Contributions of single-cell recording studies. *Ciba Foundation Symposium*, *107*, 64–82.

Deniau, J.-M., Kitai, S., Donoghue, J., & Grofova, I. (1982). Neuronal interactions in the substancia nigra pars reticulata through axon collateral of the projection neurons. *Experimental Brain Research*, *47*, 105–113.

Dupuis, P., & Nagurney, A. (1993). Dynamical systems and variational inequalities. *Annals of Operations Research*, *44*(1), 7–42.

Filippov, A. F. (1988). *Differential equations with discontinuous righthand sides*. Kluwer Academic Pub.

Frank, M., Santamaria, A., O'Reilly, R., & Willcutt, E. (2007). Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology*, *32*, 1583–1599.

Gerkey, B., Vaughan, R., & Howard, A. (2003). The player/stage project: Tools for multi-robot and distributed sensor systems. In *11th International conference on advanced robotics* (pp. 317–323).

Gillies, A., & Arbruthnott, G. (2000). Computational models of the basal ganglia. *Movement Disorders*, *15*(5), 762–770.

Girard, B., Cuzin, V., Guillot, A., Gurney, K. N., & Prescott, T. J. (2003). A basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of Integrative Neuroscience*, *2*(2), 179–200.

Girard, B., Filliat, D., Meyer, J.-A., Berthoz, A., & Guillot, A. (2005). Integration of navigation and action selection in a computational model of cortico-basal ganglia-thalamo-cortical loops. *Adaptive Behavior*, *13*(2), 115–130.

Girard, B., Tabareau, N., Berthoz, A., & Slotine, J.-J. (2006). Selective amplification using a contracting model of the basal ganglia. In F. Alexandre, Y. Boniface, L. Bougrain, B. Girau, & N. Rougier (Eds.), *NeuroComp* (pp. 30–33).

Girard, B., Tabareau, N., Slotine, J.-J., & Berthoz, A. (2005). Contracting model of the basal ganglia. In J. Bryson, T. Prescott, & A. Seth (Eds.), *Modelling natural action selection: Proceedings of an international workshop* (pp. 69–76). Brighton, UK: AISB Press.

Gurney, K., Humphries, M., Wood, R., Prescott, T., & Redgrave, P. (2004). Testing computational hypotheses of brain systems function: A case study with the basal ganglia. *Network: Computation in Neural Systems*, *15*, 263–290.

Gurney, K., Prescott, T., Wickens, J., & Redgrave, P. (2004). Computational models of the basal ganglia: From membranes to robots. *Trends in Neurosciences*, *27*, 453–459.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, *84*, 401–410.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biological Cybernetics*, *84*, 411–423.

Horn, R., & Johnson, C. (1985). *Matrix analysis*. Cambridge University Press.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–271). Cambridge, MA: The MIT Press.

Humphries, M., Gurney, K., & Prescott, T. (2005). Is there an integrative center in the vertebrate brain-stem? A robotic evaluation of a model of the reticular formation viewed as an action selection device. *Adaptive Behavior*, *13*(2), 97–113.

Ioannou, P., & Sun, J. (1996). *Robust adaptive control*. Upper Saddle River, NJ, USA: Prentice Hall, Inc.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, *15*(4–6).

Juraska, J., Wilson, C., & Groves, P. (1977). The substancia nigra of the rat: A golgi study. *Journal of Comparative Neurology*, *172*, 585–600.

Khamassi, M., Girard, B., Berthoz, A., & Guillot, A. (2004). Comparing three critic models of reinforcement learning in the basal ganglia connected to a detailed actor part in a s-r task. In F. Groen, N. Amato, A. Bonarini, E. Yoshida, & B. Krse (Eds.), *Proceedings of the eighth international conference on intelligent autonomous systems*, IAS8. (pp. 430–437). Amsterdam, The Netherlands: IOS Press.

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. *Adaptive Behavior*, *13*(2), 131–148.

Kimura, A., & Graybiel, A. (Eds.) (1995). *Functions of the cortico-basal ganglia loop*. Tokyo, New York: Springer.

Kita, H., Tokuno, H., & Nambu, A. (1999). Monkey globus pallidus external segment neurons projecting to the neostriatum. *Neuroreport*, *10*(7), 1476–1472.

Konidaris, G., & Barto, A. (2006). An adaptive robot motivational system. In S. Nolfi, G. Baldassarre, R. Calabretta, J. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, & D. Parisi (Eds.), *LNAI: Vol. 4095. From animals to animats 9: Proceedings of the 9th international conference on the simulation of adaptive behavior* (pp. 346–356). Berlin, Germany: Springer.

Krotopov, J., & Etlinger, S. (1999). Selection of actions in the basal ganglia thalamocortical circuits: Review and model. *International Journal of Psychophysiology*, *31*, 197–217.

Lohmiller, W., & Slotine, J. (1998). Contraction analysis for nonlinear systems. *Automatica*, *34*(6), 683–696.

Lohmiller, W., & Slotine, J. (2000). Nonlinear process control using contraction analysis. *American Institute of Chemical Engineers Journal*, *46*(3), 588–596.

Middleton, F. A., & Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, *266*, 458–461.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, *50*(4), 381–425.

Montes-Gonzalez, F., Prescott, T. J., Gurney, K. N., Humphries, M., & Redgrave, P. (2000). An embodied model of action selection mechanisms in the vertebrate brain. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, & S. W. Wilson (Eds.), *From animals to animats 6: Vol. 1* (pp. 157–166). Cambridge, MA: The MIT Press.

Parent, A., Sato, F., Wu, Y., Gauthier, J., Lévesque, M., & Parent, M. (2000). Organization of the basal ganglia: The importance of the axonal collateralization. *Trends in Neuroscience*, *23*(10), S20–S27.

Park, M., Falls, W., & Kitai, S. (1982). An intracellular HRP study of rat globus pallidus. I. responses and light microscopic analysis. *Journal of Comparative Neurology*, *211*, 284–294.

Pirjanian, P. (1999). *Behavior coordination mechanisms — state-of-the-art. Technical report IRIS-99-375*. Institute of Robotics and Intelligent Systems, School of Engineering, University of Southern California.

Prescott, T. J., Montes-Gonzalez, F., Gurney, K., Humphries, M. D., & Redgrave, P. (2006). A robot model of the basal ganglia: Behavior and intrinsic processing. *Neural Networks*, *19*, 31–61.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience*, *89*(4), 1009–1023.

Sato, F., Lavallee, P., Lévesque, M., & Parent, A. (2000). Single-axon tracing study of neurons of the external segment of the globus pallidus in primates. *Journal of Comparative Neurology*, *417*, 17–31.

Slotine, J., & Coetsee, J. (1986). Adaptive sliding controller synthesis for nonlinear systems. *International Journal of Control*, *43*(4), 1631–1651.

Slotine, J. J. E., & Lohmiller, W. (2001). Modularity, evolution, and the binding problem: A view from stability theory. *Neural networks*, *14*(2), 137–145.

Staines, W., Atmadja, S., & Fibiger, H. (1981). Demonstration of a pallidostriatal pathway by retrograde transport of HRP-labelled lectin. *Brain Research*, *206*, 446–450.

Tabareau, N., & Slotine, J. (2006). Notes on contraction theory. Arxiv preprint nlin.AO/0601011.

Tepper, J., & Bolam, J. (2004). Functional density and specificity of neostrital interneurons. *Current Opinion in Neurobiology*, *14*, 685–692.

Tepper, J., Koós, T., & Wilson, C. (2004). Gabaergic microcircuits in the neostriatum. *Trends in Neuroscience*, *11*, 662–669.

Tyrrell, T. (1993). The use of hierarchies for action selection. *Adaptive Behavior*, *1*(4), 387–420.

Wu, Y., Richard, S., & Parent, A. (2000). The organization of the striatal output system: A single-cell juxtacellular labeling study in the rat. *Neuroscience Research*, *38*, 49–62.

Zhang, D., & Nagurney, A. (1995). On the stability of projected dynamical systems. *Journal of Optimization Theory and Applications*, *85*(1), 97–124.

6.2 (Khamassi et al, 2005)

Adaptive **Behavior**

Actor–Critic Models of Reinforcement Learning in the Basal Ganglia: From Natural to Artificial Rats

Mehdi Khamassi[1,2], Loïc Lachèze[1], Benoît Girard[1,2], Alain Berthoz[2], Agnès Guillot[1]
[1]*AnimatLab, LIP6, Paris, France*
[2]*LPPA, CNRS–Collège de France, Paris, France*

Since 1995, numerous Actor–Critic architectures for reinforcement learning have been proposed as models of dopamine-like reinforcement learning mechanisms in the rat's basal ganglia. However, these models were usually tested in different tasks, and it is then difficult to compare their efficiency for an autonomous animat. We present here the comparison of four architectures in an animat as it performs the same reward-seeking task. This will illustrate the consequences of different hypotheses about the management of different Actor sub-modules and Critic units, and their more or less autonomously determined coordination. We show that the classical method of coordination of modules by mixture of experts, depending on each module's performance, did not allow solving our task. Then we address the question of which principle should be applied efficiently to combine these units. Improvements for Critic modeling and accuracy of Actor–Critic models for a natural task are finally discussed in the perspective of our *Psikharpax* project—an artificial rat having to survive autonomously in unpredictable environments.

Keywords animat approach · TD learning · Actor–Critic model · S–R task · taxon navigation

1 Introduction

This work aims at adding learning capabilities in the architecture of action selection introduced by Girard, Filliat, Meyer, Berthoz, and Guillot (2005) in this issue. This architecture will be implemented in the artificial rat *Psikharpax*, a robot that will exhibit at least some of the capacities of autonomy and adaptation that characterize its natural counterpart (Filliat et al., 2004). This learning process capitalizes on Actor–Critic architectures, which have been proposed as models of dopamine-like reinforcement learning mechanisms in the rat's basal ganglia (Houk, Adams, & Barto, 1995). In such models, an Actor network learns to select actions in order to maximize the weighted sum of future rewards, as computed on line by another network, a Critic. The Critic predicts this sum by comparing its estimation of the reward with the actual one by means of a temporal difference (TD) learning rule, in which the error between two successive predictions is used to update the synaptic weights (Sutton & Barto, 1998). A recent review of numerous computational models, built on this principle since 1995, highlighted several issues raised by the inconsistency of the detailed implementation of Actor and Critic modules with known basal ganglia anatomy and physiology (Joel, Niv, & Ruppin, 2002). In the first section of this paper, we will consider some of the main issues,

Correspondence to: Mehdi Khamassi, AnimatLab, LIP6, 8 rue du capitaine Scott, 75015 Paris, France.

(2005), Vol 13(2): 131–148.
[1059–7123(200506) 13:2; 131–148; 054250]

updated with anatomical and neurophysiological knowledge. In the second section, we will illustrate the consequences of alternative hypotheses concerning the various Actor–Critic designs by comparing animats that perform the same classical instrumental learning (S–R task). During the test, the animat freely moves in a plus-maze with a reward placed at the end of one arm. The reward site is chosen randomly at the beginning of each trial and it refers to site-specific local stimuli. The animat has to autonomously learn to associate continuous sensory information with certain values of reward and to select sequences of behaviors that enable it to reach the goal from any place in the maze. This experiment is more realistic than others used to validate Actor–Critic models, often characterized by an a priori fixed temporal interval between a stimulus and a reward (e.g., Suri & Schultz, 1998), by an unchanged reward location over trials (e.g., Strösslin, 2004), or by a discrete state space (e.g., Baldassarre, 2002).

We will compare, in this task, four different principles inspired by Actor–Critic models trying to tackle the issues evoked in the first section. The first one is the seminal model proposed by Houk et al. (1995), which uses one Actor and a single prediction unit (model AC: One Actor, one Critic), which is supposed to induce learning in the whole environment. The second principle implements one Actor with several Critics (model AMC1: One Actor, multiple Critics). The Critics are combined by a mixture of experts where a gating network is used to decide which expert—which Critic—is used in each region of the environment, depending on its performance in that region. The principle of mixture of experts is inspired from several existing models (Jacobs, Jordan, Nowlan, & Hinton, 1991; Baldassarre, 2002; Doya, Samejima, Katagiri, & Kawato, 2002). The third one is inspired by Suri and Schultz (2001) and also uses one Actor with several Critic experts. However, the decision of which expert should work in each sub-zone of the environment is independent of the experts' performances, but rather depends on a partition of the sensory space perceived by the animat (model AMC2: One Actor, multiple Critics). The fourth one (model MAMC2: Multiple Actors, multiple Critics) proposes the same principle as the previous Critic, combined with several Actors, which latter principle is one of the features of the model of Doya et al. (2002), particularly designed for continuous tasks, and is also a feature of Baldassarre's model (2002). Here we implement these principles in four models using the same design for each Actor component. A comparison is made of the learning speed and of their ability to extend learning to the whole experimental environment.

The last section of the paper discusses the results on the basis of acquired knowledge in reinforcement learning tasks in artificial and natural rodents.

2 Actor–Critic Designs: The Issues

The two main principles of Actor–Critic models that lead them to be considered as a good representation of the role of the basal ganglia in reinforcement learning of motor behaviors are (i) the implementation of a TD learning rule which leads to progressive translation of reinforcement signals from the time of reward occurrence to environmental contexts that precede the reward, and (ii) the separation of the model into two distinct parts: One for the selection of motor behaviors (actions) depending on the current sensory inputs (the Actor), and the other for the driving of the learning process via dopamine signals (the Critic).

Schultz's work on the electrophysiology of dopamine neurons in monkeys showed that dopamine patterns of release are similar to the TD learning rule (see Schultz, 1998 for a review). Besides, the basal ganglia are a major input to dopamine neurons, and are also a privileged target of reinforcement signals sent by these neurons (Gerfen, Herkenham, & Thibault, 1987). Moreover, the basal ganglia appears to be comprised of two distinct sub-systems, related to two different parts of the striatum—the major input nucleus of the basal ganglia—one projecting to motor areas in the thalamus, the other projecting to dopamine neurons, influencing the firing patterns of these neurons at least to some extent (Joel & Weiner, 2000).

These properties lead the first Actor–Critic model of the basal ganglia to propose the matrisomes of the striatum to constitute the Actor, and the striosomes of this very structure to be the Critic (Houk et al., 1995, Figure 1). The classical segregation of "direct" and "indirect" pathways from the striatum to the dopaminergic system (SNc, substantia nigra pars compacta, and VTA, ventral tegmental area; Albin, Young, & Penney, 1989) was used in the model to explain the timing characteristics of dopamine neurons' discharges.

Numerous models have been proposed to improve and complete the model of Houk et al. However, most

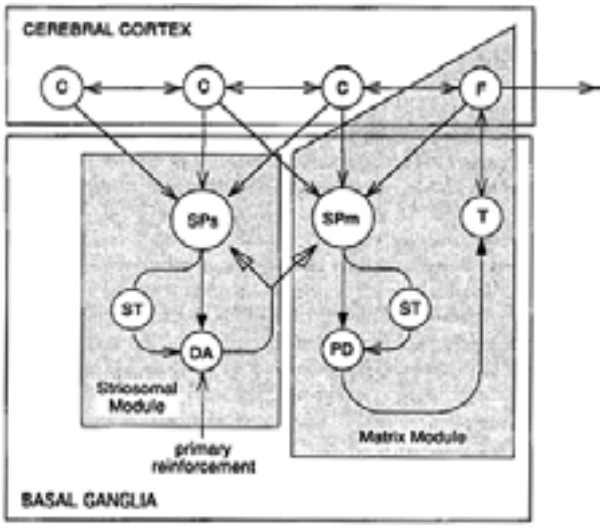

Figure 1 Schematic illustration of the correspondence between the modular organization of the basal ganglia including both striosomes and matrix modules and the Actor–Critic architecture in the model proposed by Houk et al. (1995). F: columns in the frontal cortex; C: other cortical columns; SPs: spiny neurons striosomal compartments of the striatum; SPm: spiny neurons in matrix modules; ST: subthalamic sideloop; DA: dopamine neurons in the substantia nigra compacta; PD: pallidal neurons; T: thalamic neurons. (Adapted from Houk et al., 1995.)

of these computational models have neurobiological inconsistencies and incompleteness concerning recent anatomical hypotheses on the basal ganglia (Joel et al., 2002).

An important drawback is that the Actor part of these models is often simplistic compared to the known anatomy of the basal ganglia and does not take into account important anatomical and physiological characteristics of the striatum. For example, recent works showed a distinction between neurons in the striatum having different dopamine receptors (D1-receptors or D2-receptors; Aizman et al., 2000). This implies at least two different pathways in the Actor, on which tonic dopamine has opposite effects, going beyond the classical functional segregation of "direct" and "indirect" pathways in the striatum (Gurney, Prescott, & Redgrave, 2001a,b).

Likewise, some constraints deriving from striatal anatomy restrict the possible architectures for the Critic network. In particular, the striatum is constituted of only one layer of medium spiny neurons—completed with 5% of interneurons (Houk et al., 1995). As a consequence, Critic models cannot be constituted of complex multilayer networks for reward prediction computation. This anatomical constraint led several authors to model the Critic as a single-neuron (Houk et al., 1995; Montague, Dayan, & Sejnowski, 1996), which works well in relatively simple tasks. For more complicated tasks, several models assign one single Critic neuron to each subpart of the task. These models differ in the computational mechanism used to coordinate these neurons. Baldassarre (2002) and Doya et al. (2002) propose to coordinate Critic modules with a mixture of experts method: The module that has the best performance at a certain time during the task becomes expert in the learning process of this subpart of the task. Another model proposes an association of experts with subparts of the task (such as stimuli or events) in an a priori manner, independently from each expert's performance (Suri & Schultz, 2001). It remains to assess the efficiency of each principle, as they have been at work in heterogeneous tasks (e.g., Wisconsin Card Sorting Test, Discrete Navigation Task, Instrumental Conditioning).

These models also question the functional segregation of the basal ganglia in "direct" and "indirect" pathways (see Joel et al., 2002 for a review). These objections are built on electrophysiological data (for a review see Bunney, Chiodo, & Grace, 1991) and anatomical data (Joel & Weiner, 2000) which show that these two pathways are unable to produce the temporal dynamics necessary to explain dopamine neurons' patterns of discharge. These findings lead one to question the localization of the Critic in the striosomes of the dorsal striatum, and several models have capitalized on its implementation in the ventral striatum (Brown, Bullock, & Grossberg, 1999; Daw, 2003). These works are supported by recent fMRI data in humans, showing a functional dissociation between dorsal striatum as the Actor and ventral striatum as the Critic (O'Doherty et al., 2004), but they may be controversial for the rat, as electrophysiological data (Thierry, Gioanni, Dégénetais, & Glowinski, 2000) showed that an important part of the ventral striatum (the nucleus accumbens core) does not project extensively to the dopamine system in the rat brain.

We can conclude that the precise implementation of the Critic remains an open question, if one takes also into account a recent model assuming that a new functional distinction of striosomes in the dorsal striatum—based on differential projections to GABA-A

and GABA-B receptors in dopamine neurons—can explain the temporal dynamics expected (Frank, Loughry, & O'Reilly, 2001).

Besides these neurobiological inconsistencies, some computational requirements on which numerous Actor–Critic models have focused seem unnecessary for a natural reward-seeking task. For example, as Houk et al.'s model could not account for temporal characteristics of dopamine neurons firing patterns, most of the alternative models focused on the simulation of the depression of dopamine at the precise time where the reward is expected when it eventually does not occur. To this purpose, they concentrated on the implementation of a temporal component for stimulus description—which is computed outside of the model and is sent as an input to the model via cortical projections (Montague et al., 1996; Schultz, Dayan, & Montague, 1997). These models were tested in the same tasks chosen by Schultz, Apicella, and Ljungberg (1993) to record dopamine neurons in the monkey, using a fixed temporal bin between a stimulus and a reward. However, in natural situations where a rodent needs to find food or any other type of reward, temporal characteristics of the task are rarely fixed but rather depend on the animal's behavior and on the environment's changes/evolution.

3 Method

The objective of this work is to evaluate the efficiency of the main principles on which current Actor–Critic models inspired by the basal ganglia are designed, when they are implemented in the same autonomous artificial system. The main addressed issues are the following:

- The implementation of a detailed Actor, whose structure would be closer to the anatomy of the dorsal striatum, assessing whether reinforcement learning is still possible within this architecture.
- The comparison of the function of one Critic unit, versus several alternative ways to coordinate different Critic modules for solving a complex task where a single-neuron is not enough.
- The test of the models in a natural task involving taxon navigation where events are not predetermined by fixed temporal bins. Instead, the animat perceives a continuous sensory flow during its movements, and has to reactively switch its actions so as to reach a reward.

3.1 The Simulated Environment and Task

Figure 2 shows the experimental setup simulated, consisting in a simple 2D plus-maze. The dimensions are equivalent to a 5 m × 5 m environment with 1-m large corridors. In this environment, walls are made of segments colored on a 256 grayscale. The effects of lighting conditions are not simulated. Every wall of the maze is colored in black (luminance = 0), except walls at the end of each arm and at the center of the maze, which are represented by specific colors: The cross at the center is gray (191), three of the arm extremities' walls are dark gray (127) and the fourth is white (255), indicating the reward location (equivalent to a water trough delivering two drops—noninstantaneous reward—not a priori known by the animat).

The plus-maze task mimics the neurobiological and behavioral studies that will serve as future validation for the model (Albertin, Mulder, Tabuchi, Zugaro, & Wiener, 2000). In this task, at the beginning of each trial, one arm extremity is randomly chosen to deliver reward. The associated wall is colored in white whereas walls at the three other extremities are dark gray. The animat has to learn that selecting the action *drinking* when it is near the white wall (distance < 30 cm) and faces it (angle < 45°) gives it a reward. Here we assume that reward = 1 for n iterations ($n = 2$), without considering how the hedonic value of this reward is determined.

We expect the animat to learn a sequence of context-specific behaviors, so that it can reach the reward site from any starting point in the maze:

- When not seeing the white wall, face the center of the maze and move forward.
- As soon as arriving at the center (the animat can see the white wall), turn to the white stimulus.
- Move forward until being close enough to reward location.
- Drink.

The trial ends when reward is consumed: The color of the wall at reward location is changed to dark gray, and a new arm extremity is chosen randomly to deliver reward. The animat has then to perform again

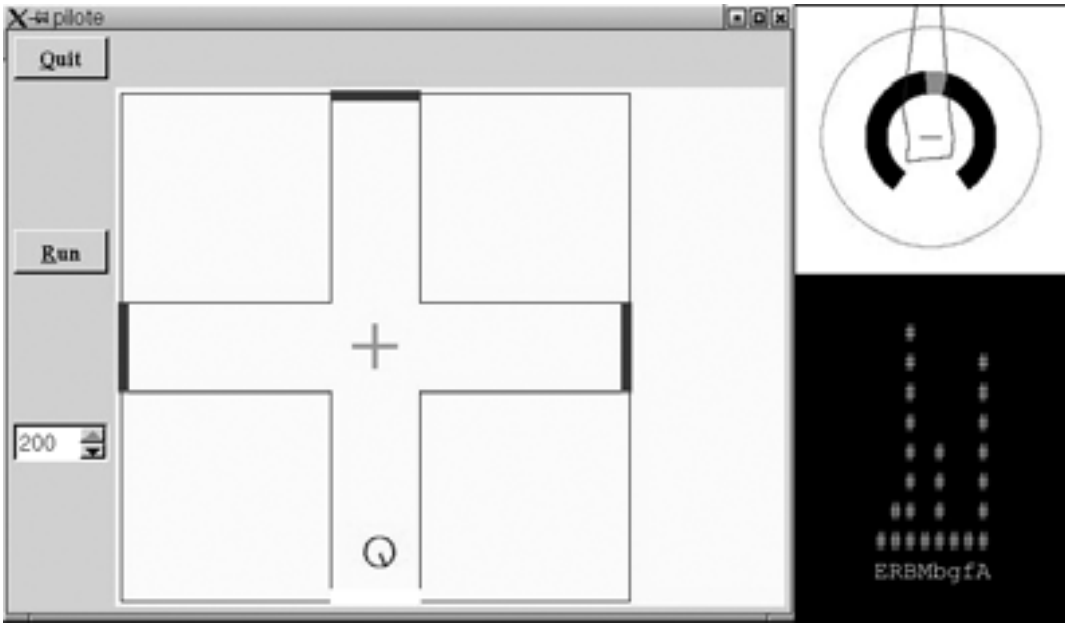

Figure 2 Left: the robot in the plus-maze environment. A white arm extremity indicates the reward location. Other arm extremities do not deliver any reward and are shown in black. Upper right: the robot's visual perceptions. Lower right: activation level of different channels in the model.

the learned behavioral sequence. Note that there is no break between two consecutive trials: Trials follow each other successively.

The more efficiently and fluently the animat performs the above-described behavioral sequence, the less time it will take to reach the reward. As a consequence, the criterion chosen to validate the models is the time to goal, plotted along the experiment as the learning curve of the model.

3.2 The Animat

The animat is represented by a circle (30-cm diameter). Its translation and rotation speeds are 40 cm s^{-1} and 10° s^{-1}. Its simulated sensors areas follows:

- an omnidirectional linear camera providing at every 10° the color of the nearest perceived segment; this results in a 36-color table that constitutes the animat's visual perception (see Figure 2);
- eight sonars with a 5-m range, an incertitude of ± 5° concerning the pointed direction and an additional ± 10-cm measurement error.

The sonars are used by a low-level obstacle avoidance reflex which overrides any decision taken by the Actor–Critic model when the animat comes too close to obstacles.

The animat is provided with a visual system that computes 12 input variables ($\forall i \in [1;12], 0 < \text{var}_i < 1$) out of the 36-color table at each time step. These sensory variables constitute the state space of the Actor–Critic and so will be taken as input to both the Actor and the Critic parts of the model (Figure 3). Variables are computed as follows:

- *seeWhite* (resp. *seeGray*, *seeDarkGray*) = 1 if the color table contains the value 255 (resp. 191, 127), else 0.
- *angleWhite*, *angleGray*, *angleDarkGray* = (number of boxes in the color table between the animat's head direction and the desired color)/18.
- *distanceWhite*,*distanceGray, distanceDarkGray* = (maximum number of consecutive boxes in the color table containing the desired color)/18.
- *nearWhite* (resp. *nearGray*, *nearDarkGray*) = 1 – *distanceWhite* (resp. *distanceGray, distanceDarkGray*).

Representing the environment with such continuous variables implies the model permanently receiving a flow of sensory information and having to learn

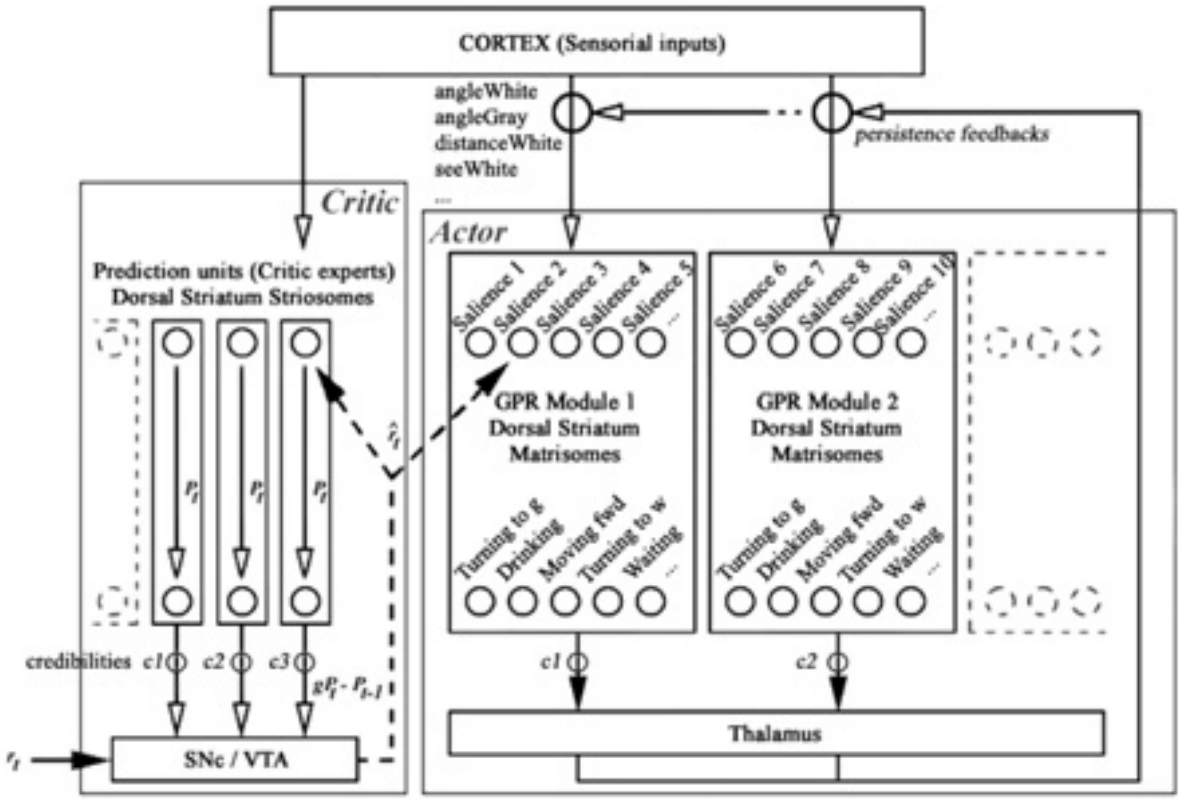

Figure 3 General scheme of the models tested in this work. The Actor is a group of GPR modules with saliences as inputs and actions as outputs. The Critic (involving striosomes in the dorsal striatum, and the substantia nigra compacta (SNc)) propagates towards the Actor an estimate $\check{r}$ of the instantaneous reinforcement triggered by the selected action. The particularity of this scheme is to combine several modules for both Actor and Critic, and to weight the Critic experts' predictions and the Actor modules' decisions with credibilities. These credibilities can be either computed by a gating network (model AMC1) or in a context-dependent manner (models AMC2 and MAMC2).

autonomously the events (sensory contexts) that can be relevant for the task resolution.

The animat has a repertoire of 6 actions: *Drinking*, *moving forward*, *turning to white perception*, *turning to gray perception*, *turning to dark gray perception*, and *waiting*. These actions constitute the output of the Actor model (described below) and the input to a low-level model that translates it into appropriate orders to the animat's engines.

3.3 The Model: Description of the Actor Part

The Actor–Critic model is inspired by the rat basal ganglia. As mentioned in Section 2, the Actor can be hypothesized as implemented in the matrix part of the basal ganglia, while striosomes in the dorsal striatum are considered as the anatomical counterpart for the Critic. The Critic produces dopamine-like reinforcement signals that help it learn to predict reward during the task, and that make the Actor learn to select appropriate behaviors in every sensory context experienced during the task.

The architecture implemented in the Actor is a recent model proposed by Gurney, Prescott, and Redgrave (2001a,b)—henceforth called the GPR model—that replaces the simple winner-takes-all which usually constitutes Actor models and is supposed to be more biologically plausible.

Like other Actors, the GPR model consists of a series of parallel channels, each one representing an action (in our implementation, we used 6 channels corresponding to the 6 actions used for the task). This architecture constitutes an alternative view to the prevailing functional segregation of the basal ganglia into "direct" and "indirect" pathways discussed in Section 1

(Gurney et al., 2001a,b). All these channels are composed of two different circuits through the dorsal striatum: The first is the "selection" pathway, implementing action selection properly via a feed-forward off-center on-surround network, and mediated by cells in the dorsal striatum with D1-type receptors. The second is the "control" pathway, mediated by cells with D2-type receptors in the same area. Its role is to regulate the selection by enhancing the selectivity inter-channels, and to control the global activity within the Actor. Moreover, a cortex–basal-ganglia–thalamus loop in the model allows it to take into account each channel's persistence in the process of selection (see Gurney et al., 2001a,b, for detailed description and mathematical implementation of the model). The latter characteristic showed some interesting properties that prevented a robot from performing behavioral oscillations (Montes-Gonzalez, Prescott, Gurney, Humphries, & Redgrave, 2000; Girard, Cuzin, Guillot, Gurney, & Prescott, 2003).

In our implementation, the input values of the Actor model are saliences—i.e., the strength of a given action—that are computed out of the 12 sensory variables, a constant implementing a bias, and a persistence factor—equal to 1 for the action that was selected at previous timestep (Figure 3). At each timestep t (timesteps being separated by a 1-s bin in our simulations), the action that has the highest salience is selected to be performed by the animat, the salience of action i being

$$\mathrm{sal}_i(t) = \left[\sum_{j-1}^{13} \mathrm{var}_j(t) \cdot w_{i,j}(t)\right] + \mathrm{persist}_i(t) \cdot w_{i,14}(t) \quad (1)$$

where $\mathrm{var}_{13}(t) = 1$, $\forall t$, and the $w_{i,j}(t)$ are the synaptic weights representing, for each action i, the association strength with input variable j. These weights are initiated randomly ($\forall i, j, -0.02 < w_{i,j}(t = 0) < 0.02$) and the objective of the learning process will be to find a set of weights allowing the animat to perform the task efficiently.

An exploration function is added that would allow the animat to try an action in a given context even if the weights of the Actor do not give a sufficient tendency to perform this action in the considered context.

To do so, we introduce a clock that triggers exploration in two different cases:

- When the animat has been stuck for a large number of timesteps (*time* superior to a fixed threshold α) in a situation that is evaluated negative by the model (when the prediction $P(t)$ of reward computed by the Critic is inferior to a fixed threshold);
- When the animat has remained for a long time in a situation where $P(t)$ is high but this prediction does not increase that much ($|P(t + n) - P(t)| < \varepsilon$) and no reward occurs.

If one of these two conditions is true, exploration is triggered: One of the 6 actions is chosen randomly. Its salience is set to 1 (note that when exploration = false, $\mathrm{sal}_i(t) < 1$, $\forall i, t, w_{i,j}(t)$) and is maintained at 1 for a duration of 15 timesteps (the time necessary for the animat to make a 180° turn or to run from the center of the maze to the end of one arm).

3.4 The Model: Description of the Critic Part

For the Critic part of the model, different principles based on existing techniques are tested. The idea is to test the hypothesis of one single Critic unit first, but also to provide the Critic with enough computational capacities so that it can correctly estimate the value function over the whole environment of the task. In other words, the Critic will have to deal with several different sensory contexts—corridors, maze center, extremity of arms, etc., equivalent to different stimuli—and will have to associate a correct reward prediction to these contexts.

One obvious possibility would be a multilayer perceptron with several hidden layers but, as mentioned in Section 2, there are anatomical constraints which prevent us from adopting this choice: Our Critic is supposed to be situated in the striosomes of dorsal striatum, which structure is constituted of only one layer of medium spiny neurons (Houk et al., 1995). Thus we need a more general method that combines several Critic modules, each one being constituted of a single neuron and dealing with a particular part of the problem space.

The method adopted here is the mixture of experts, which was proposed to divide a nonlinearly separable problem into a set of linearly separable problems, and to affect a different expert to each considered sub-problem (Jacobs, Jordan, Nowlan, & Hinton, 1991).

The Critics tested in this work differ mainly in the two following manners:

- the first (model AMC1) implements a mixture of experts in which a gating network is used to decide which expert is used in each region;
- the second (model AMC2) implements a mixture of experts in which a hand-determined partition of the environment based on a categorization of visual perceptions is used to decide which expert works in each subzone.

Moreover, since the animat has to solve a task in continuous state space, there could be interferences between reinforcement signals sent by different Critic experts to the same single Actor. In this way, whereas one model will employ only one Actor (model AMC2), another one will use one Actor module associated to each expert (model MAMC2). Figure 3 shows the general scheme with different modules employed as suggested by the models presented here.

Performances of models AMC1, AMC2 and MAMC2 will be compared, together with the one of the seminal Actor–Critic model inspired by the basal ganglia, proposed by Houk et al. (1995), and using a single cell Critic with a single Actor (model AC).

We start with the description of the simplest Critic, the one belonging to model AC.

3.4.1 Model AC In this model, at each timestep, the Critic is a single linear cell that computes a prediction of reward based on the same input variables as the Actor, except for the persistence variable

$$P(t) = \sum_{j=1}^{13} \mathrm{var}_j(t) \cdot w'_j(t) \tag{2}$$

where $w'_j(t)$ are the synaptic weights of the Critic.

This prediction is then used to calculate the reinforcement signal by means of the TD-rule:

$$\hat{r}(t) = r(t) + gP(t) - P(t-1) \tag{3}$$

where $r(t)$ is the actual reward received by the animat, and g is the discount factor ($0 < g < 1$) which determines how far in the future expected rewards are taken into account in the sum of future rewards.

Finally, this reinforcement signal is used to update both Actor's and Critic's synaptic weights according to the following equations respectively:

$$w_{i,j}(t) \leftarrow w_{i,j}(t-1) + \eta \cdot \hat{r}(t) \cdot \mathrm{var}_j(t-1) \tag{4}$$

$$w'_j(t) \leftarrow w'_j(t-1) + \eta \cdot \hat{r}(t) \cdot \mathrm{var}_j(t-1) \tag{5}$$

where $\eta > 0$ is the learning rate.

3.4.2 Model AMC1 As this Critic implements N experts, each expert k computes its own prediction of reward at timestep t:

$$p_k(t) = \sum_{j=1}^{13} w'_{k,j}(t) \cdot \mathrm{var}_j(t) \tag{6}$$

where the $w'_{k,j}(t)$ are the synaptic weights of expert k.

Then the global prediction of the Critic is a weighted sum of experts' predictions:

$$P(t) = \sum_{k=1}^{N} \mathrm{cred}_k(t) \cdot p_k(t) \tag{7}$$

where $\mathrm{cred}_k(t)$ is the credibility of expert k at timestep t. These credibilities are computed by a gating network which learns to associate, in each sensory context, the best credibility with the expert that makes the smaller prediction error. Following Baldassarre's description (2002), the gating network is constituted of N linear cells which receive the same input variables than the experts and compute an output function out of it:

$$o_k(t) = \sum_{j=1}^{13} w''_{k,j}(t) \cdot \mathrm{var}_j(t) \tag{8}$$

where $w''_{k,j}(t)$ are the synaptic weights of gating cell k.

The credibility of expert k is then computed as the softmax activation function of the outputs $o_f(t)$:

$$\mathrm{cred}_k(t) = \frac{o_k(t)}{\sum_{f=1}^{N} o_f(t)}. \tag{9}$$

Concerning learning rules, whereas Equation 3 is used to determine the global reinforcement signal sent to

the Actor, each Critic's expert has a specific reinforcement signal based on its own prediction error:

$$\hat{r}_k(t) = r(t) + gP(t) - p_k(t-1)\,. \tag{10}$$

The synaptic weights of each expert k are updated according to the following formula:

$$w''_{k,j}(t) \leftarrow w''_{k,j}(t-1) + \eta \cdot \hat{r}_k(t) \cdot var_j(t-1) \cdot h_k(t) \tag{11}$$

where $h_k(t)$ is the contribution of expert k to the global prediction error of the Critic, and is defined as

$$h_k(t) = \frac{\text{cred}_k(t-1) \cdot \text{corr}_k(t)}{\sum_{f=1}^{N} \text{cred}_f(t-1) \cdot \text{corr}_f(t)} \tag{12}$$

where $\text{corr}_k(t)$ is a measure of the correctness of the expert k defined as

$$\text{corr}_k(t) = \exp\left(\frac{-\hat{r}_k(t)^2}{2\sigma^2}\right) \tag{13}$$

where σ is a scaling parameter depending on the average error of the experts (see table of parameters in the Appendix).

Finally, to update the weights of the gating network, we use the following equation:

$$w''_{k,j}(t) \leftarrow w''_{k,j}(t-1) + m \cdot \text{diff}(t) \cdot \text{var}_j(t-1) \tag{14}$$

with $\text{diff}(t) = h_k(t) - \text{cred}_k(t-1)$ where m is a learning rate specific to the gating network.

So the credibility of expert k in a given sensory context depends on its performance in this context.

3.4.3 Model AMC2 This Critic also implements N experts. However, it differs from model AMC1 in the way the credibility of each expert is computed.

The principle we want to bring about here is to dissociate credibilities of experts from their performance. Instead, experts are assigned to different subregions of the environment (these regions being computed as windows in the perceptual space) remain enchained to their associate region forever, and progressively learn to improve the accuracy of their performance during the experiment. This principle is adopted from Houk et al. (1995) for the improvement of their model, assuming that different striosomes may be specialized in dealing with different behavioral tasks. This proposition was implemented by Suri and Schultz (2001) in using several TD models, each one computing predictions for only one event (stimulus or reward) that occurs in the simulated paradigm.

To test this principle, we replaced the gating network by a hand-determined partition of the environment (e.g., a coarse representation of the sensory space): At timestep t, the current zone β depends on the 12 sensory variables computed by the visual system. Example: If (*seeWhite* = 1 *and angleWhite* < 0.2 *and distanceWhite* > 0.8) then *zone* = 4 (e.g., $\beta = 4$). Then $\text{cred}_\beta(t) = 1$, $\text{cred}_k(t) = 0$ for all other experts, and expert β has then to compute a prediction of reward out of the 12 continuous sensory variables. Predictions and reinforcement signals of the experts are determined by the same equations as Critic of model AMC1.

This was done as a first step in the test of the considered principle. Indeed, we assume that another brain region such as the parietal cortex or the hippocampus would determine the zone (sensory configuration) depending on the current sensory perception (McNaughton, 1989; Burgess, Jeffery, & O'Keefe, 1999), and would send it to the Actor–Critic model of the basal ganglia. Here, the environment was partitioned into $N = 30$ zones, an expert being associated with each zone. The main difference between this scheme and the one used by Suri and Schultz is that, in their work, training of experts in each sub-zone was done in separated sessions, and the global model was tested on the whole task only after training of all experts. Here, experts are trained simultaneously in a single experiment.

Finally, one should note that this method is different from applying a coarse coding of the state space that constitutes the input to the Actor and the Critic (Arleo & Gerstner, 2000). Here, we implemented a *coarse coding of the credibility space* so as to determine which expert is the most credible in a given sensory configuration, and kept the 12 continuous sensory variables, plus a constant described above, as the state space for the reinforcement learning process. This means that within a given zone, the concerned expert has to learn to approximate a con-

tinuous reward value function, based on the varying input variables.

3.4.4 Model MAMC2 The Critic of this model is the same as in model AMC2 and only differs in its associated Actor.

Instead of using one single Actor, we implemented N different Actor modules. Each Actor module has the same structure as the simple Actor described in Section 3.4 and consists of six channels representing the six possible actions for the task. The difference resides in the fact that only actions of the Actor associated with the zone in which the animat is currently are competing to determine the animat's current action.

As a consequence, if the animat is in zone β at time t and performed action i, the reinforcement signal $\hat{r}(t+1)$ computed by the Critic at next timestep will be used to update only weights of action i from the Actor β according to the following equation:

$$w_{k,i,j}(t) \leftarrow w_{k,i,j}(t-1) + \eta \cdot \hat{r}(t) \cdot \mathrm{var}_j(t-1). \quad (15)$$

Other equations are the same as those used for Critic of model AMC2. As mentioned above, this principle (using a specific controller or a specific Actor for each module of the Actor–Critic model) is inspired by the work of Doya et al. (2002).

3.5 Results

In order to compare the learning curves of the four simulated models, and so as to evaluate which models manage to solve the task efficiently, we adopt the following criterion: After 50 trials of training (out of 100 for each experiments), the animat has to achieve an equivalent performance to a hand-crafted model that can already solve the task (Table 1). To do so, we simulated the GPR action selection model with appropriate hand-determined synaptic weights and without any learning process, so that the animat can solve the task as if it had already learned it. With this model, the animat performed a 50-trial experiment with an average performance of 142 iterations per trial. Since each iteration lasted approximately 1 s, as mentioned above, it took a little bit more than 2 min per trial for this hand-crafted animat to reach the reward.

Table 1 shows the performance of each model, measured as the average number of iterations per trial after trial #50. Figure 4 illustrates results to the four experiments performed in the 2D environment, one per model. The x-axis represents the successive trials during the experiments. For each trial, the y-axis shows the number of iterations needed for the animat to reach the reward and consume it. Figure 4a shows the learning curve of model AC. It can be seen that the model rapidly increased its performance until trial 7, and stabilized it at trial 25. However, after trial 50, the average duration of a trial is still 587 iterations, which is nearly 4 times higher than the chosen criterion. We can explain this limitation by the fact that model AC consists of only one single neuron in the Critic, which can only solve linearly separable problems. As a consequence, the model could learn only a part of the task (in the area near the reward location), and was unable to extend learning to the rest of the maze. So the animat has learned to select appropriate behaviors in the reward area, but it still performs random behaviors in the rest of the environment.

Table 1 Performance of each model.

Model	GPR	AC	AMC1	AMC2	MAMC2
Performance	142	587	623	3240	97

Model AMC1 is designed to mitigate the computational limitations of model AC, as it implies several Critic units controlled by a gating network. Figure 4b shows its learning curve after simulation in the plus-maze task. The model has also managed to decrease its running time per trial at the beginning of the experiment. However, it can be seen that the learning process is more unstable than the previous one. Furthermore, after the 50th trial, the model has a performance of 623 iterations, which is no better than model AC. Indeed, the model could not extend learning to the whole maze either. We can explain this failure by the fact that the gating network did not manage to specialize different experts in different subparts of the task. As an example, Figure 5 shows the reward prediction computed by each Critic's expert during the last trial of the experiment. It can be noticed that the first expert (dark curve) has the highest prediction throughout the whole trial. This is due to the fact that it is the only one the gating network has learned to consider as credible—its credibility remains above 90% during the whole experiment. As a consequence, only one expert is involved in the learning process and the model becomes computa-

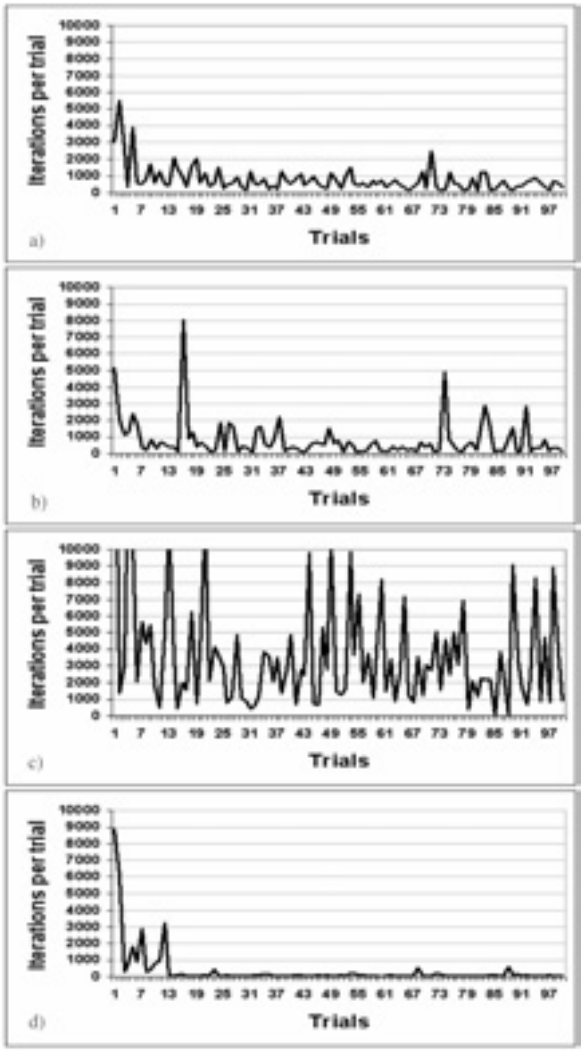

Figure 4 Learning curves of the four models simulated in the 2D plus-maze task over 100 trials experiments: *x*-axis, trials; *y*-axis, number of iterations per trial (truncated to 10000 for better readability). (a) Model AC, (b) model AMC1, (c) model AMC2, (d) model MAMC2.

tionally equivalent to model AC: It cannot extend learning to the whole maze, which is confirmed by the absence of any reward prediction before the perception of the reward site (stimulus occurrence) in Figure 5.

Figure 4c shows the learning curve of model AMC2 which implements another principle for experts coordination. This model does not suffer from the same limitations as model AMC1, since each expert was a priori assigned to a specific area of the environment. As a consequence, it quickly managed to extend learning to the whole maze. However, the consequence of

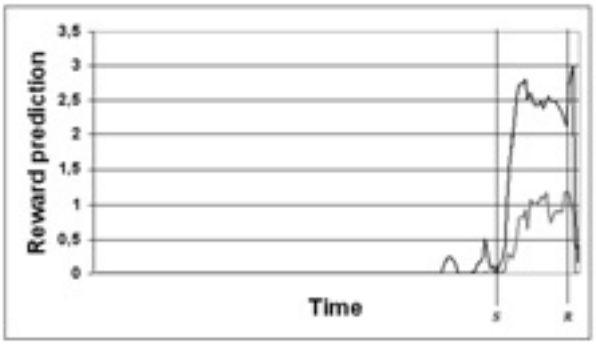

Figure 5 Reward prediction computed by each Critic's expert of model AMC1 during trial #100 of the experiment. Time 0 indicates the beginning of the trial. S: perception of the stimulus (the white wall) by the animat. R: beginning of reward delivery. The dark curve represents the prediction of expert 1. The other experts' predictions are melted into the light curve or equal to 0.

this process is to produce interferences in the Actor's computations: The same Actor receives all experts' teaching signals, and it remains unable to switch properly between reinforced behaviors. For example, when the action drinking is reinforced, the Actor starts selecting this action permanently, even when the animat is far from reward location. These interferences explain the very bad performances obtained with model AMC2.

The last simulated model (model MAMC2) performed best. Its learning curve is shown in Figure 4d. This model implements several Actor modules (an Actor module connected to each Critic expert). As a consequence, it avoids interferences in the learning process and rapidly converged to a performance of 97 iterations per trial. This good performance cannot be reached with the multi-Actor only; we tried to combine several Actor modules to model AMC1 and got a performance of 576 iterations per trial. So the achievement of the task implies a combination of multi-Actor and a good specialization of experts.

To check the ability of model MAMC2 to learn the same task in more realistic conditions, we simulated it in a 3D environment, working in real time and implementing physical dynamics (Figure 7). This experiment involved an intermediary step favoring the implementation into an actual Pekee robot (Wany Robotics). The animat is still able to learn the task in this environment and gets good performances after 35 trials (Figure 6; corresponding average performance

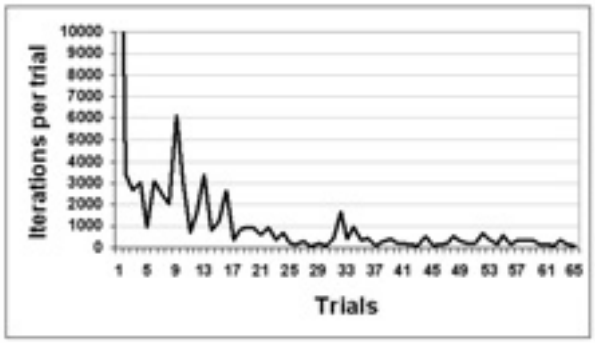

Figure 6 Learning curve in the 3D environment: *x*-axis, trials; *y*-axis, number of iterations per trial.

of the animat between trials 35 and 65: 284 iterations per trial).

4 Discussion and Future Work

In this work, we have compared learning capabilities on a S–R task of several Actor–Critic models of the basal ganglia based on distinct principles. Results of simulations with models AC, AMC1, AMC2 and MAMC2 demonstrated that

- a single-component Critic cannot solve the task (model AC);
- several Critic modules controlled by a gating network (model AMC1) cannot provide good specialization, and the task remains unsolved;
- several Critic modules a priori associated with different subparts of the task (model AMC2) and connected to a single Actor (an Actor component being composed of a 6-channel GPR) allow learning to extend to areas that are distant from reward location, but still suffer from interferences between signals sent by the different Critic to the same single Actor.

Model MAMC2, combining several Critic modules with the principle of model AMC2, and implementing several Actor components, produces better results in the task, spreading learning in the whole maze and reducing the learning duration. However, there are a few

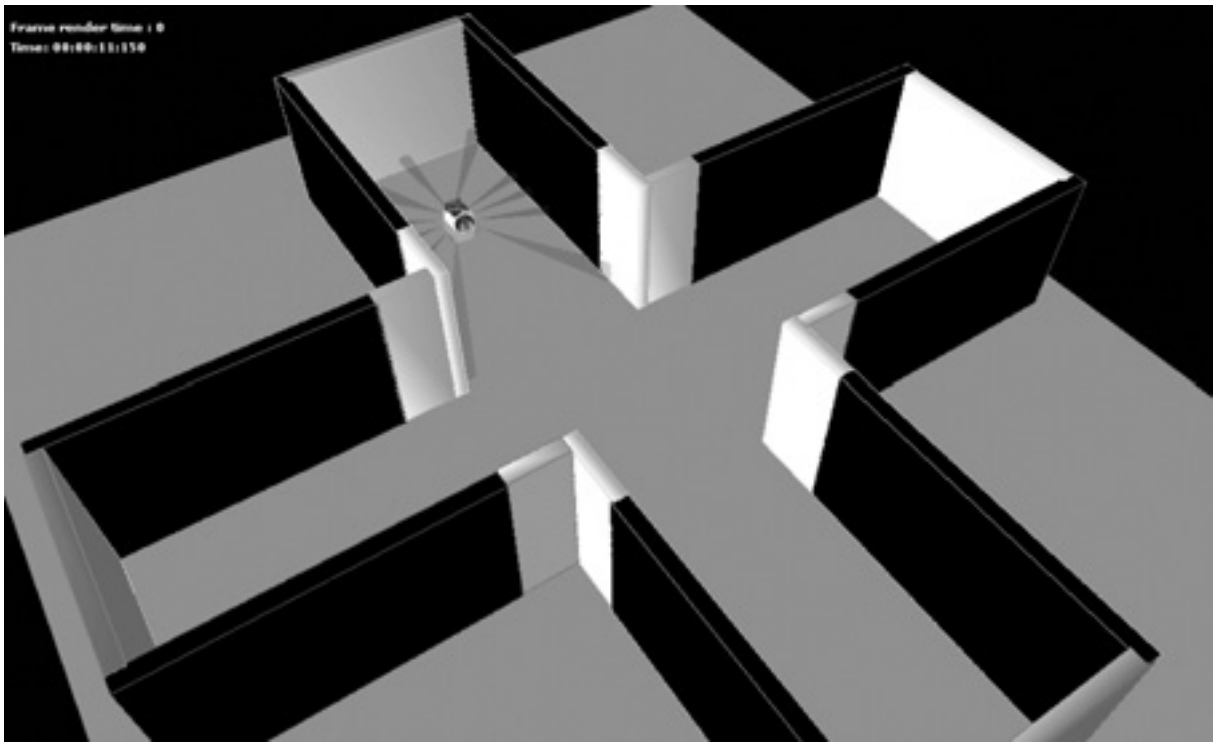

Figure 7 Simulation of the plus-maze task in a 3D environment. Like the 2D environment, one random arm extremity is white and delivers reward. The animat has to perform taxon navigation so as to find and consume this reward. Gray stripes arising from the animat's body represent its sonar sensors used by its low level obstacle avoidance reflex.

questions that have to be raised concerning the biological plausibility and the generalization ability of this model.

4.1 Biological Plausibility of the Proposed Model

When using a single GPR Actor, each action is represented in only one channel—an Actor module consisting of one channel per action (Gurney et al., 2001a,b)—and the structural credit assignment problem (which action to reinforce when getting a reward) can be simply solved: The action that has the highest salience inhibits its neighbors via local recurrent inhibitory circuits within D1 striatum (Brown & Sharp, 1995). As a consequence, only one channel in the Actor will have enough pre- and post-synaptic activity to be eligible for reinforcement.

When using several Actor modules, this property is no longer true: Even if only one channel per Actor module may be activated at a given time, each Actor module will have its own activated channel, and several concurring synapses would be eligible for reinforcement within the global Actor. To solve this problem, we considered in our work that only one channel in the entire Actor is eligible at a given time. However, this implies that the basal ganglia has one of the two following characteristics: Either there should exist non-local inhibition between Actor modules within the striatum, or there should be some kind of selectivity in the dopamine reinforcement signals so that even if several channels are activated, only those located in the target module receive dopamine signals.

To the best of our knowledge, these characteristics have not been found in the basal ganglia, and some studies tend to refute the dopamine selectivity (Pennartz, 1996).

4.2 Computational Issues

Several computational issues need also to be addressed. First, the results presented here show that the learning process was not perturbed by the fact to use an Actor detailing the action selection process in the basal ganglia. This Actor has the property to take into account some persistence provided by the cortex–basal-ganglia–thalamus–cortex loops. The way this persistence precisely influence the learning process in the different principles compared in this work was not thoroughly studied here. However, we suspect that persistence could probably challenge the way different Actors interact with Critic's experts, as switching between actions does not exactly follow switches in sensorimotor contexts with this model. This issue should be examined in a future work.

4.2.1 Generalization ability of the multi-module Actor. Another issue that needs to be addressed here is the generalization ability of the multi-module Actor model used in this experiment. Indeed, model MAMC2 avoids interferences in the Actor because hand-determined subzones of the maze are absolutely disjoint. In other words, learned stimulus–response associations in a given zone cannot be performed in another zone, and do not interfere with the learning process is this second zone even if visual contexts associated with each of them are very similar. However, this also leads to an inability to generalize from one zone to the other: Even if the distinction we made between two zones seemed relevant for the plus-maze task, if these two zones were similar and implied similar motor responses in another task, the animat would have to learn the same sensorimotor association twice—one time in each zone. As a consequence, the partition we set in this work is task-dependent.

Alternatively, the model would need a partitioning method that autonomously classifies sensory contexts independently from the task, can detect similarities between two different contexts and can generalize learned behaviors from the first experienced context to the second one.

4.2.2 About the precise time of reward delivery. In the work presented here, the time of reward delivery depends exclusively on the animat's behavior, which differs from several other S–R tasks used to validate Actor–Critic models of the basal ganglia. In these tasks, there is a constant duration between a stimulus and a reward, and several Actor–Critic models have been designed to describe the precise temporal dynamics of dopaminergic neurons in this type of task (Montague et al., 1996). As a consequence, numerous Actor–Critic models focused on the implementation of a time component for stimulus representation, and several works capitalized on this temporal repre-

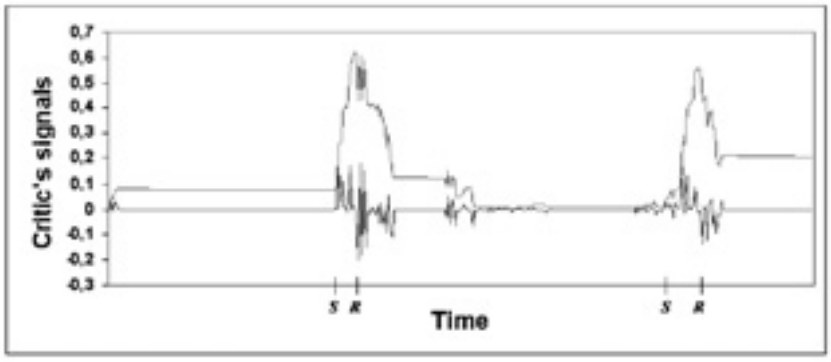

Figure 8 Reward prediction (light curve) and dopamine reinforcement signal (dark curve) computed by Critic of model MAMC2 in the 3D environment: *x* -axis, time; *y*-axis, Critic's signals. S: perception of the stimulus (white wall) by the animat; R: Reward missed by the animat.

sentation for the application of Actor–Critic models of reinforcement learning in the basal ganglia to robotics (Perez-Uribe, 2001; Sporns & Alexander, 2002). Will we need to add such a component to our model to be able to apply it to a certain type of natural task, or survival task?

In the experiments presented here, we did not need such a temporal representation of stimuli because there was sufficient information in the continuous sensory flow perceived by the animat during its moves, so that the model could dynamically adapt its reward predictions, as observed also by Baldassarre and Parisi (2000). For example, when the animat is at the center of the maze, perceives the white wall (stimulus predicting reward) and moves towards reward location, the latter stimulus becomes bigger in the visual field of the animat, and the model can learn to increase its reward prediction, as shown in Figure 8. We did not aim to explain the depression of dopamine neurons' firing rates when a reward does not occur; nevertheless, we were able to observe this phenomenon in cases where the animat was approaching the reward site, was about to consume it, but finally turned away from it (R events in Figure 8).

4.2.3 Using Critics dependent or independent from the performance. In our experiments, model AMC1, implementing a gating network for experts' credibilities computation, did not solve the task. We saw in Section 2 that, during the simulations, one expert became rapidly the most credible, which forced the model to use only one neuron to solve the task. The use of gating networks in the frame of mixture of experts methods has already been criticized (Tang, Heywood, & Shepherd, 2002). According to these authors, this approach works well on problems composed of disjoint regions but does not generalize well, suffering from effects on boundaries of regions.

In our case, we explain the failure in the experts' specialization with model AMC1 by the observation that until the model has started to learn the task, and so can propagate teaching signals to the rest of the maze, only reward location has a value. As a consequence, it is the only area where the gating network tries to train an expert, and the latter rapidly reaches a high credibility. Then, as reward value starts to be extended to a new zone, this same expert still has the best credibility while getting bad performances. Other experts do not have significantly better performances—since they were not trained yet and since the new area and the first one are not disjoint. As a consequence, they remain noncredible and the model starts having bad performances.

Baldassarre (2002) managed to obtain a good specialization of experts. This may be partly explained by the fact that his task involved three different rewards located in three different sensory contexts. The simulated robot had to visit all rewards alternately from the very beginning of the task. This may have helped the gating network to attribute good credibilities to several experts. However, reward locations in Baldassarre's task are not perfectly disjoint, which results in a difficult specialization: One of the experts is the most credible for two of the three rewards (see Baldassarre, 2002).

Another model (Tani & Nolfi, 1999) proposes a different mixture of experts where the gating network

is replaced with a dynamical computation of experts' credibilities. Their model managed to categorize the sensory–motor flow perceived by a simulated robot during its movements. However, their method does not use any memory of associations between experts' credibilities and different contexts experienced during the task. As a consequence, experts' specialization is even more dependent on each expert's performance than Baldassarre's gating network, and suffers from the same limitation when applied to reinforcement learning in our plus-maze task—as we have found in experiment (unpublished work).

4.2.4 Combining self-organizing maps with mixture of expert. To test the principle of dissociating the experts credibility from their performance, we partitioned the environment into several sub-regions. However, this method is ad hoc, lacks autonomy, and suffers generalization abilities if the environment is changed or becomes more complex. We are currently implementing self-organizing maps (SOMs) as a method of autonomous clustering of the different sensory contexts that will be used to determine these zones. Note that this proposition differs from the traditional use of SOMs to cluster the state space input to experts or to Actor–Critic models (Smith, 2002; Lee & Kim, 2003). It is rather a clustering of the credibility space, which was recently proposed by Tang et al. (2002). We would also like to compare the use of SOMs with the use of place cells. Indeed models of hippocampal place cells have already been used for coarse coding of the input state space to the Actor and the Critic (Arleo & Gerstner, 2000; Foster, Morris, & Dayan, 2000; Strösslin, 2004) but, in our case, we would like to use place cells to determine experts' credibilities.

4.3 Future Work

As often mentioned in the literature, and as confirmed in this work, the application of Actor–Critic architectures to continuous tasks is more difficult than their use in discrete tasks. Several other works have been done on the subject (Doya, 2000). However, these architectures still have to be improved so as to decrease their learning time.

Particularly, the learning performance of our animat seems still far from the learning speed that real rat can reach in the same task (Albertin et al., 2000), even if the high time constant that we used in our model does not allow a rigorous comparison yet (see the table of parameters in the Appendix). This could be at least partly explained by the fact that we implemented only S–R learning (or habit learning), whereas it has recently been known that rats are endowed with two distinct learning systems related to different cortex–basal-ganglia–thalamus loops: A habit learning system and a goal-directed learning one (Ikemoto & Panksepp, 1999; Cardinal, Parkinson, Hall, & Everitt, 2002). The latter would be fast, used at the early stages of learning, and implies an explicit representation of rewarding goals or an internal representation of action-outcome contingencies. The former would be very slow and takes advantage of the latter when the animat achieves good performance and becomes able to solve the task with a reactive strategy (S–R) (Killcross & Coutureau, 2003; Yin, Knowlton, & Balleine, 2004).

Some theoretical work has already been started to extend Actor–Critic models to this functional distinction (Dayan, 2001). In the practical case of our artificial rat, both such systems could be useful in two different manners.

First, it could be useful to upgrade the exploration function. This function could have an explicit representation of different places of the environment, and particularly of the reward site. Then, when the animat gets reward for the first time, the exploration function would guide it, trying behaviors that can allow it to reach the explicitly memorized reward location. The function could also remember which behaviors have already been tried unsuccessfully in the different areas, so that untried behaviors are selected instead of random behaviors in the case of exploration. This would strengthen the exploration process and is expected to increase the animat's learning speed.

The second possible use of a goal-directed behavior component is to represent the type of reward the animat is working for. This can be useful when an animat has to deal with different rewards (food, drink) so as to satisfy different motivations (hunger, thirst). In this case, a component that chooses explicitly the current reward the animat takes as an objective can select sub-modules of the Actor that are dedicated to the sequence of behaviors that leads to the considered reward. This improvement would serve as a more realistic validation of the artificial rat *Psikharpax* when it has to survive in more natural environments, satisfying concurrent motivations.

Appendix

Table 2 Parameters.

Symbol	Value	Description
Δt	1 s	Time constant: Time between two successive iterations of the model.
α	40 iterations	Time threshold to trigger the exploration function.
g	0.98	Discount factor of the temporal difference learning rule.
η	0.01	Learning rate of the Actor and Critic modules.
N	30	Number of experts in the Critic of models AMC1, AMC2 and MAMC2.
σ	2	Scaling parameter in the mixture of experts of model AMC1.
m	0.1	Learning rate of the gating network in model AMC1.

Acknowledgments

This research has been supported by the LIP6 and the Project *Robotics and Artificial Entities* (ROBEA) of the Centre National de la Recherche Scientifique, France. Thanks for useful discussions go to Angelo Arleo, Gianluca Baldassarre, Francesco Battaglia, Etienne Koechlin and Jun Tani.

References

Aizman, O., Brismar, H., Uhlen, P., Zettergren, E., Levey, A. I., Forssberg, H., Greengard, P., & Aperia, A. (2000). Anatomical and physiological evidence for D1 and D2 dopamine receptors colocalization in neostriatal neurons. *Nature Neuroscience*, *3*(3), 226–230.

Albertin, S. V., Mulder, A. B., Tabuchi, E., Zugaro, M. B., & Wiener, S. I. (2000). Lesions of the medial shell of the nucleus accumbens impair rats in finding larger rewards, but spare reward-seeking behavior. *Behavioral Brain Research*, *117*(1–2), 173–183.

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neuroscience*, *12*, 366–375.

Arleo, A., & Gerstner, W. (2000). Spatial cognition and neuromimetic navigation: A model of the rat hippocampal place cell activity. *Biological Cybernetics*, Special Issue on Navigation in Biological and Artificial Systems, *83*, 287–299.

Baldassarre, G. (2002). A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviors. *Journal of Cognitive Systems Research*, *3*(1), 5–13.

Baldassarre, G., & Parisi, D. (2000). Classical and instrumental conditioning: From laboratory phenomena to integrated mechanisms for adaptation. In J.-A. Meyer, A. Berthoz, D. Floreana, H. L. Roitblat, and S. W. Wilson (Eds.), *From animals to animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior,* supplement volume (pp. 131–139). Cambridge, MA: The MIT Press.

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning, or incentive salience? *Brain Research Reviews*, *28*, 309–369.

Brown, L., & Sharp. F. (1995). Metabolic mapping of rat striatum: Somatotopic organization of sensorimotor activity. *Brain Research*, *686*, 207–222.

Bunney, B. S., Chiodo, L. A., & Grace, A. A. (1991). Midbrain dopamine system electrophysiological functioning: A review and new hypothesis. *Synapse*, *9*, 79–84.

Burgess, N., Jeffery, K. J., & O'Keefe, J. (1999). Integrating hippocampal and parietal functions: A spatial point of view. In N. Burgess, K. J. Jeffery, and J. O'Keefe (Eds.), *The hippocampal and parietal foundations of spatial cognition* (pp. 3–29). Oxford: Oxford University Press.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum and prefrontal cortex. *Neuroscience Biobehavioral Reviews*, *26*(3), 321–352.

Dayan., P. (2001). Motivated reinforcement learning. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Proceedings of NIPS 14* (pp. 11–18). Cambridge, MA: The MIT Press.

Daw, N. D. (2003). *Reinforcement learning models of the dopamine system and their behavioral implications*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Doya, K. (2000). Reinforcement learning in continuous time and space. *Neural Computation*, *12*, 219–245.

Doya, K., Samejima, K., Katagiri, K., & Kawato, M. (2002) Multiple model-based reinforcement learning. *Neural Computation*, *14*(6), 1347–1369.

Filliat, D., Girard, B., Guillot, A., Khamassi, M., Lachèze, L., & Meyer, J.-A. (2004). State of the artificial rat Psikharpax. In S. Schaal, A. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam, and J.-A. Meyer (Eds.), *From animals to animats 8: Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior* (pp. 2-12). Cambridge, MA: The MIT Press.

Foster, D., Morris, R., & Dayan, P. (2000). Models of hippocampally dependent navigation using the temporal difference learning rule. *Hippocampus*, *10*, 1–16.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective and Behavioral Neuroscience*, *1*(2), 137–160.

Gerfen, C. R., Herkenham, M., & Thibault, J. (1987). The neostriatal mosaic: II. Patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. *Journal of Neuroscience*, *7*, 3915–3934.

Girard, B., Cuzin, V., Guillot, A., Gurney, K., & Prescott, T. (2003). A basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of Integrative Neuroscience*, *2*(22), 179–200.

Girard, B., Filliat, D., Meyer, J.-A., Berthoz, A., & Guillot, A. (2005). Integration of navigation and action selection functionalities in a computational model of cortico-basal-thalamo-cortical loops. *Adaptive Behavior*, 13 (2), 115–130.

Gurney, K. N., Prescott, T. J., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia: I. A new functional anatomy. *Biological Cybernetics*, *84*, 401–410.

Gurney, K. N., Prescott, T. J., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia: II. Analysis and simulation of behavior. *Biological Cybernetics*, *84*, 411–423.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, and D. G. Beiser (Eds.), *Models of information processing in the basal ganglia*, Cambridge, MA: The MIT Press.

Ikemoto, S., & Panksepp., J. (1999). The role of the nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, *31*, 6–41.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive Mixture of Local Experts. *Neural Computation*, *3*, 79–87.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, *15*, 535–547.

Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with striatum in rats and primates: An analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, *96*, 451–474.

Killcross, A. S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex*, *13*(4), 400–408.

Lee, J. K., & Kim, I. H. (2003). Reinforcement learning control using self-organizing map and multi-layer feed-forward neural network. In *Proceedings of the International Conference on Control Automation and Systems, ICCAS 2003* (pp. 142–145). Gyeongju, South Korea.

McNaughton, B. L. (1989). Neural mechanisms for spatial computation and information storage. In L. Nadel, L. A. Cooper, P. Harnish, and R. M. Colicover (Eds.), *Neural Connections, Mental Computations* (chapter 9, pp. 285-350). Cambridge, MA: MIT Press.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Montes-Gonzalez, F. Prescott, T. J., Gurney, K. N., Humphries, M., & Redgrave, P. (2000). An embodied model of action selection mechanisms in the vertebrate brain. In J.-A. Meyer, A. Bethoz, D. Floreana, H. L. Roitblat, and S. W. Wilson (Eds.), *From animals to animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior* (pp.157–166). Cambridge, MA: The MIT Press.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. (2004). Dissociable roles of dorsal and ventral striatum in instrumental conditioning. *Science*, *304*, 452–454.

Pennartz, C. M. A. (1996). The ascending neuromodulatory systems in learning by reinforcement: Comparing computational conjectures with experimental findings. *Brain Research Reviews*, *21*, 219–245.

Perez-Uribe, A. (2001). Using a time-delay actor–critic neural architecture with dopamine-like reinforcement signal for learning in autonomous robots. In S. Wermter J. Austin, and D. Willshaw (Eds.), *Emergent neural computational architectures based on neuroscience: A state-of-the-art survey* (pp. 522–533). Berlin: Springer.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, *13*(3), 900–913.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Smith, A. J. (2002). Applications of the self-organizing map to reinforcement learning. *Neural Networks*, *15*(8-9), 1107–1124.

Sporns, O., & Alexander, W. H. (2002). Neuromodulation and plasticity in an autonomous robot. *Neural Networks*, *15*, 761–774.

Strösslin, T. (2004). *A connectionist model of spatial learning in the rat.* Ph.D thesis, EPFL, Swiss Federal Institute of Technology.

Suri, R. E., & Schultz, W. (2001). Temporal difference model reproduces anticipatory neural activity. *Neural Computation*, *13*, 841–862.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: The MIT Press.

Tang, B., Heywood, M. I., & Shepherd, M. (2002). Input partitioning to mixture of experts. In IEEE/INNS *International Joint Conference on Neural Networks* (pp. 227–232), Honolulu, Hawaii (pp. 227–232).

Tani, J., & Nolfi, S. (1999). Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. *Neural Networks*, *12*(7-8), 1131–1141.

Thierry, A.-M., Gioanni, Y., Dégénetais, E., & Glowinski, J. (2000). Hippocampo-prefrontal cortex pathway: Anatomical and electrophysiological characteristics. *Hippocampus*, *10*, 411–419.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*(1), 181–189.

About the Authors

Mehdi Khamassi is working as a Ph.D. student in cognitive science both at the AnimatLab of the Laboratoire d'Informatique de Paris 6 (LIP6) and at the Laboratoire de Physiologie de la Perception et de l'Action (LPPA–CNRS, College de France). He trained as an engineer and received a master's degree in cognitive science from the University Pierre and Marie Curie (UPMC Paris 6). His current research interests include electrophysiology experiments and computational modeling of learning processes in the rat brain.

Loïc Lachèze is working at the AnimatLab as a Ph.D. student. He received a master's degree in computer science from the University of Paris 6 in 2002. He currently contributes to the *Psikharpax* project and works on the robotic integration of visual process and control architectures of navigation and action selection. *Address*: AnimatLab, LIP6, 8 rue du capitaine Scott, 75015 Paris, France. E-mail: loic.lacheze@lip6.fr

Benoît Girard was trained as an engineer at the Ecole Centrale de Nantes (ECN), he received a Ph.D. in computer science (2003) from the University Pierre and Marie Curie (UPMC Paris 6). He is now a Post-Doc Fellow at the Laboratoire de Physiologie de la Perception et de l'Action (LPPA–CNRS, College de France) where he works on models of the primate saccadic circuitry. His research is focused on biomimetic neural network models of navigation, action selection and motor execution. *Address*: LPPA, Collège de France, 11 place Marcellin Berthelot, 75005 Paris, France. E-mail: benoit.girard@college-de-france.fr

Alain Berthoz was trained as an engineer. He graduated in human psychology and received a Ph.D. in biology. He is Professor at the Collège de France, where he heads the Laboratoire de Physiologie de la Perception et de l'Action (LPPA). He is a member of numerous academic societies, an invited expert in several international committees, and he has been awarded with many prizes. His main scientific interests are in the multisensory control of gaze, of equilibrium, of locomotion and of spatial memory. He coordinates the neurophysiological experiments that inspire the *Psikharpax* project. *Address*: LPPA, Collège de France, 11 place Marcellin Berthelot, 75005 Paris, France. E-mail: alain.berthoz@college-de-france.fr

Agnès Guillot is Associate Professor of Psychophysiology at the University of Paris X. She graduated in human and animal psychology, and holds Ph.D.s in psychophysiology and biomathematics. Her main scientific interests are in action selection in animals and robots. She coordinates the biomimetic modeling of *Psikharpax* at the AnimatLab of the Laboratoire d'Informatique de Paris 6 (LIP6). *Address:* AnimatLab, LIP6, 8 rue du capitaine Scott, 75015 Paris, France. E-mail: agnes.guillot@lip6.fr

6.3 (Colas et al, 2009)

Biol Cybern (2009) 100:203–214
DOI 10.1007/s00422-009-0292-y

ORIGINAL PAPER

Biological Cybernetics

Bayesian models of eye movement selection with retinotopic maps

Francis Colas · Fabien Flacher · Thomas Tanner · Pierre Bessière · Benoît Girard

Received: 31 July 2008 / Accepted: 9 January 2009 / Published online: 11 February 2009

Abstract Among the various possible criteria guiding eye movement selection, we investigate the role of position uncertainty in the peripheral visual field. In particular, we suggest that, in everyday life situations of object tracking, eye movement selection probably includes a principle of reduction of uncertainty. To evaluate this hypothesis, we confront the movement predictions of computational models with human results from a psychophysical task. This task is a freely moving eye version of the multiple object tracking task, where the eye movements may be used to compensate for low peripheral resolution. We design several Bayesian models of eye movement selection with increasing complexity, whose layered structures are inspired by the neurobiology of the brain areas implied in this process. Finally, we compare the relative performances of these models with regard to the prediction of the recorded human movements, and show the advantage of taking explicitly into account uncertainty for the prediction of eye movements.

Keywords Bayesian modeling · Retinotopic maps · Eye movements selection · Multiple-object tracking

F. Colas (✉) · F. Flacher · B. Girard
Laboratoire de Physiologie de la Perception et de l'Action,
CNRS/Collège de France, 11 pl. Marcelin Berthelot,
75231 Paris Cedex 05, France
e-mail: colas.francis@gmail.com

F. Flacher
e-mail: fabien.flacher@gmail.com

B. Girard
e-mail: benoit.girard@college-de-france.fr

T. Tanner
Department of Cognitive and Computational Psychophysics,
MPI for Biological Cybernetics, Spemannstr. 38,
72076 Tübingen, Germany
e-mail: tanner@tuebingen.mpg.de

P. Bessière
Laboratoire d'Informatique de Grenoble, CNRS/Grenoble Universités,
655 av. de l'Europe, 38334 Montbonnot, France
e-mail: bessiere@imag.fr

1 Introduction

We usually make a few saccades per seconds. Saccades, and other eye movements, may result from a decision on where to look next, in order to gain information about the visual scene by driving the fovea towards regions of interest. Indeed, as the sensitivity and spatial resolution of the retina decays towards the periphery of the visual field, we are uncertain about the accuracy of what we perceive in the periphery and about what we can expected to learn from an eye movement towards a peripheral position. The uncertainty is a common issue for both perception—because we cannot be sure of what we perceive—and action—because we cannot be sure of the consequences of our actions. In this paper, we investigate the possible role of uncertainty evaluation in selection processes related to active perception. We build a Bayesian model inspired by the neurophysiology of eye movement selection related brain regions, in order to investigate eye movements selection during freely moving eye multiple object tracking task (MOT).

1.1 Bayesian methodology

In order to handle uncertainty and to explicitly reason about it, we use the Bayesian Programming framework (Lebeltel et al. 2004; Bessière et al. 2008). This framework provides a systematic procedure to build and use a Bayesian model. Such a model uses probability distributions to

represent knowledge with uncertainty. It then reasons about this knowledge by applying the rules of probability theory. More precisely, starting from a joint probability distribution, marginalization and Bayes' rules allow to compute any conditional or marginal probability distribution. As this joint probability is usually of very high dimensionality, we use conditional independence hypotheses to decompose the joint distribution in a simpler product of smaller distributions.

In the end, a Bayesian programmer specifies a set of variables, a *decomposition* of the joint probability distribution and a mathematical expression for each factor that appears in this decomposition. At that point, any distribution on the variables can be computed. The programmer is usually interested on one particular distribution, which is called a *question*. The inference can be automatically computed through the use of both marginalization and Bayes rules.

1.2 Eye movement circuitry

Even if we do not have the pretension to build a complete model of the neurophysiology of the brain regions related to eye movement selection, the structure of our model is inspired by their anatomy and electrophysiology. Saccadic and smooth pursuit circuitry share a large part of their functional architecture (Krauzlis 2004). Among those regions containing saccadic and smooth pursuit subcircuits (Fig. 1), the superior colliculus (SC), the frontal eye fields (FEF) and the lateral bank of the intraparietal sulcus (LIP) in the posterior parietal cortex have a number of common points. They all receive information concerning the position of points of interest in the visual field (visual activity), memorize these positions (delay activity) and are implied in the selection of the gaze targets among these points (presaccadic activity) (Moschovakis et al. 1996; Wurtz et al. 2001; Scudder et al. 2002). These positions are encoded by cells with receptive/motor fields defined in a retinotopic reference frame. Our model is based on retinotopic probability distributions encoding similar information (observations, memory of target positions, motor decision).

In the SC, these cells are clearly organized in topographic maps, in various species (Robinson 1972; McIlwain 1976, 1983; Siminoff et al. 1966; Herrero et al. 1998). In primates, these maps have a complex logarithmic mapping (Fig. 2) (Robinson 1972; Ottes et al 1986), similar to the mapping found in the striate cortex (Schwarz 1980). Concerning the FEF, mapping studies clearly show a logarithmic encoding of the eccentricity of the position vector (Sommer and Wurtz 2000), however complementary studies are necessary to understand how its orientation is encoded. Finally, the structure of the LIP maps is still to be deciphered, even if a continuous topographical organization seems to exist, with an over representation of the central visual field (Ben Hamed et al. 2001). Given the lack of quantitatively defined FEF and

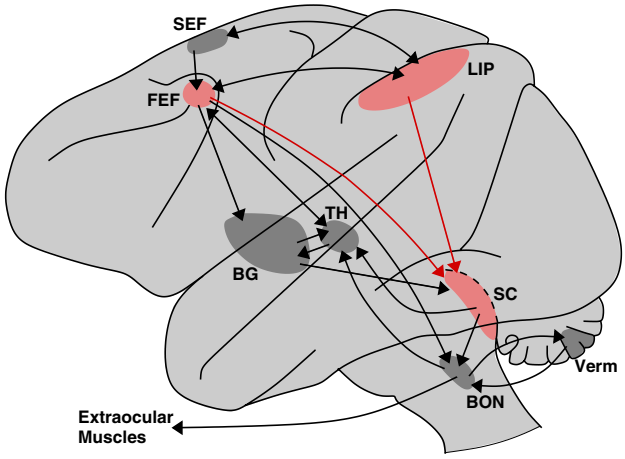

Fig. 1 Premotor and motor circuitry shared by saccade and smooth pursuit movement (Macaque monkey). *BG* basal ganglia, *BON* brainstem oculomotor nuclei, *FEF* frontal eye fields, *LIP* lateral bank of the intraparietal sulcus, *SC* superior colliculus, *SEF* supplementary eye fields, *TH* thalamus, *Verm* cerebellar vermis. In light *red* regions using retinotopic reference frames to encode visual, memory and motor activity, refer to text for more details. Adapted from (Krauzlis 2004) (color in online)

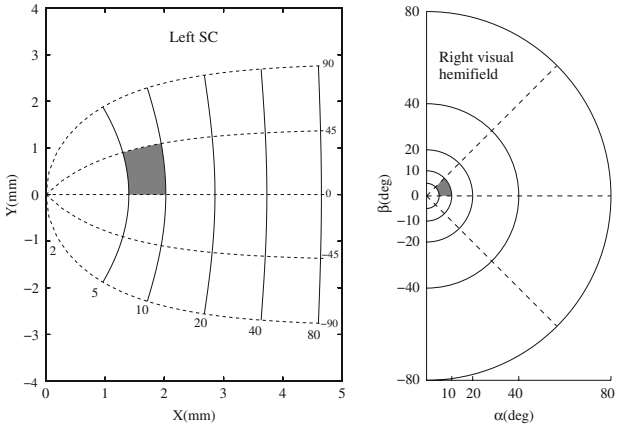

Fig. 2 Macaque collicular mapping. The angular position of targets in the visual field (*right*) are mapped onto the SC surface (*left*) using a logarithmic mapping. The *grey* areas represent the same part of the visual field in both representations

LIP mappings, we assume that they share similar properties with the SC one and thus use the log complex mapping of the SC for all the position encoding variables of our model.

The neurons related to the spatial working memory in SC (Mays and Sparks 1980), FEF (Goldberg and Bruce 1990) and LIP (Gnadt and Andersen 1988; Barash et al. 1991a,b)—also called quasi-visual cells or QV—are capable of dynamic remapping. These cells can be activated by a memory of the position of a target, even if the target was not in the cell's receptive field at the time of presentation. They behave as if they were included in a retinotopic memory map, integrating a remapping mechanism allowing the displacement of the

memorized activity when an eye movement is performed. Neural network models of that type of maps, either in the SC or the FEF, have already been proposed (Droulez and Berthoz 1991; Bozis and Moschovakis 1998; Mitchell and Zipser 2003). Such a mechanism, adapted to Bayesian programming, is used in the representation and memory layers of our model.

To summarize, though not strictly neuromimetic, the layered structure of our Bayesian model is based on log complex retinotopic maps with remapping capabilities, encoding the filtered visual input, the memorized position of targets of interests, and the generation of motor commands.

1.3 Experimental protocol

In order to study selection of eye movement in a controlled task, we use eye movement recordings from a freely moving eye version (Tanner et al. 2007) of the classical MOT task (Pylyshyn and Storm 1988). Eye movements in MOT have only recently attracted interest (Tanner et al. 2007; Fehd and Seiffert 2008; Zelinsky and Neider 2008). The original task was designed to investigate the distribution of covert attention with eye movements constrained by a fixation cross (Cavanagh and Alvarez 2005), while we looked at how free eye movements might optimize the tracking. Figure 3 illustrates this experiment in which participants are presented with a set of targets among a number of distractors. All of these objects are indiscernible 1° large discs and move in a quasi-random pattern. The task is to remember which of these objects are the targets (see Appendix A for a complete description). With this experimental paradigm, the visual scene is composed of simple geometric features therefore allowing for a study of the eye movement selection that occurs in this context.

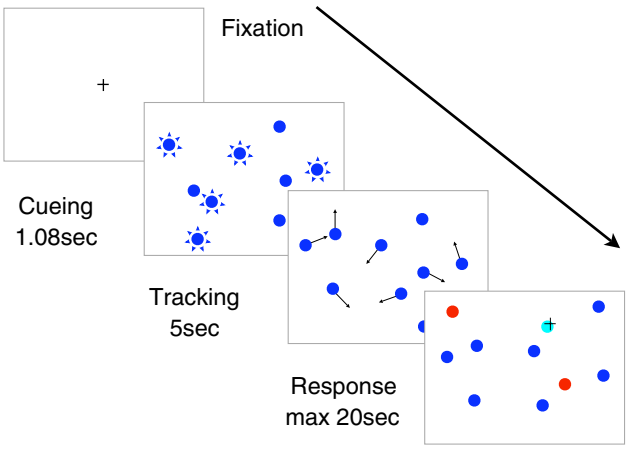

Fig. 3 Typical multiple object tracking experiment. A set of simple objects is presented, the targets are identified as the flashing ones, then the flashing stops and all the objects move around independently. After they stop moving, the subject must identify the targets

First we describe the Bayesian models we propose. Then we present the global results indicating that uncertainty is useful and some specific situations shedding light on the differences between the models.

2 Methods

The model we propose is composed of two parts. The first part deals with the perception and memory of the visual scene (*representation* model). The second part deals with the actual selection of where to look next (*decision* model).

Both models are expressed in a retinal reference frame, with a logcomplex mapping as explained above.

2.1 Representation

The representation part of our model is a dynamic retinotopic map of the visual environment. This representation is structured in two different layers. The first layer is concerned only with the integration of the visual input, i.e. the occupancy of the visual scene without any discrimination between targets and distractors (*occupancy grid*). This model would be homologous to the visual cells.

The second layer is a memory of the position of the targets, reminiscent of the QV cells. It represents the knowledge of the observer about the position of the targets, based on the occupancy representation.

2.1.1 Occupancy grid

Occupancy grids are a standard way to represent the state of an environment. They were originally introduced for the representation of obstacles in robotics applications (Elfes 1989). The general idea is to discretize the environment into a grid and to assign a variable in each cell of the grid stating whether there is an obstacle or not. The occupancy grid is therefore the collection of probability distributions over each variable in the grid.

We apply this model to the presence of objects in the visual field. More precisely, we introduce a collection O of binary variables $O^t_{(x,y)}$, one for each timestep $t \in [\![0, t_{\max}]\!]$ and location $(x, y) \in \mathcal{G}$ where $\mathcal{G}$ is a regular grid in the retino-centered logcomplex reference frame.[1] We also assume that we have visual inputs in this same reference frame, represented by a collection V of binary variables $V^t_{(x,y)}$ for $t \in [\![1, t_{\max}]\!]$ indicating if an object (either target or distractor) is perceived in the corresponding cell. Finally, we include some past eye movement information M^t in order to model the remapping

[1] Omission of an index or exponent in the variable name indicates the conjunction of all of those variables for the missing index varying in its full range: $O = O^{0\to t_{\max}} = \bigwedge_{t=0}^{t_{\max}} O^t = \bigwedge_{t=0}^{t_{\max}} \bigwedge_{(x,y)\in\mathcal{G}} O^t_{(x,y)}$.

capability exhibited by cortical and subcortical retino-centered memories.

We write the joint probability distribution over all these variables by assuming the occupancy of the cells are independent one from another conditionally to the past eye movement and the former state of the grid. We also assume that the observation corresponding to a cell is independent on all other variables conditionally to the current occupancy in this cell. This is summarized by the following factorization of the joint distribution:

$$
\begin{aligned}
&P(OVM)\\
&= P(O^0)\prod_{t=1}^{t_{\max}} P(O^t\ V^t\ M^t \mid O^{t-1})\\
&= \prod_{(x,y)\in\mathcal{G}} P(O^0_{(x,y)})\\
&\quad\times\prod_{t=1}^{t_{\max}}\left[\begin{array}{l} P(M^t)\\ \times\prod_{(x,y)\in\mathcal{G}}\left[\begin{array}{l} P(O^t_{(x,y)} \mid M^t\ O^{t-1})\\ \times P(V^t_{(x,y)} \mid O^t_{(x,y)})\end{array}\right]\end{array}\right]
\end{aligned}
$$

In this expression, $P(O^0_{(x,y)})$ is an arbitrary prior on the occupancy of the visual scene, $P(M^t)$ is a distribution over the eye movement that can be chosen arbitrarily as the results of the inference do not depend on it, provided that it is not zero for the actual eye movements observed. The relation between the occupancy and the observation, $P(V^t_{(x,y)} \mid O^t_{(x,y)})$, is a simple probability matrix chosen to state that there is a high probability of observing an object when there is one and conversely of not observing anything when there is nothing.

The evolution of the grid, with the remapping capability, is specified by the transition model, $P(O^t_{(x,y)} \mid M^t\ O^{t-1})$, which essentially transfers the probability associated to antecedent cells for the given eye movements to the corresponding present cell with an additional uncertainty factor (see Appendix B.1 for details).

With this description, updating the knowledge over the occupancy of the visual field corresponds to the following question for each time t:

$$P(O^t \mid V^{1\to t}\ M^{1\to t}) \tag{1}$$

where $V^{1\to t}$ is the conjunction of all variables V^u for $u \in [\![1, t]\!]$. This expression can be computed in an iterative manner using Bayesian inference:

$$
\begin{aligned}
&P(O^t \mid V^{1\to t}\ M^{1\to t})\\
&\quad\propto \prod_{(x,y)\in\mathcal{G}} P(V^t_{(x,y)} \mid O^t_{(x,y)})\\
&\qquad\times\sum_{O^{t-1}}\left[\begin{array}{l}\prod_{(x,y)\in\mathcal{G}} P(O^t_{(x,y)} \mid M^t\ O^{t-1})\\ \times P(O^{t-1} \mid V^{1\to t-1}\ M^{1\to t-1})\end{array}\right]
\end{aligned}
$$

However, this expression comprises a summation over all possible grid states, which is computationally intensive. Therefore we approximate the inference over the whole grid by a set of inferences for each cell that depend only on a subset of the grid:

$$
\begin{aligned}
&P(O^t_{(x,y)} \mid V^{1\to t}\ M^{1\to t})\\
&\quad\propto P(V^t_{(x,y)} \mid O^t_{(x,y)})\\
&\qquad\times\sum_{O^{t-1}_{\mathcal{A}(x,y)}}\left[\begin{array}{l} P(O^t_{(x,y)} \mid M^t\ O^{t-1}_{\mathcal{A}(x,y)})\\ \times\prod_{\mathcal{A}(x,y)} P(O^{t-1}_{(x',y')} \mid V^{1\to t-1}\ M^{1\to t-1})\end{array}\right]
\end{aligned}
$$

where $\mathcal{A}(x, y)$ is the subset of the cells (x', y') of the grid that are the antecedent of the cell (x, y) by the current eye movement M^t.

2.1.2 Positions of the targets

The previous model describes the visual scene without differentiating between targets and distractors. In order to take this two classes into account, we add a set of variables T^t_i to represent the location of each target $i \in [\![1, N]\!]$ at each time $t \in [\![0, t_{\max}]\!]$ in the logcomplex retino-centered reference frame.

This representation is the standard way to represent the location of some objects and serves a different purpose than the occupancy grid, which is only the representation of the visual scene.

The model is extended with this additional variables by adding a new factor in the joint distribution, $P(T^t_i \mid T^{t-1}_i\ O^t\ M^t)$, that represents the dynamic model of targets:

$$
\begin{aligned}
&P(OVMT)\\
&= \prod_{(x,y)\in\mathcal{G}} P(O^0_{(x,y)})\prod_{i=1}^{N} P(T^0_i)\\
&\quad\times\prod_{t=1}^{t_{\max}}\left[\begin{array}{l} P(M^t)\\ \times\prod_{(x,y)\in\mathcal{G}}\left[\begin{array}{l} P(O^t_{(x,y)} \mid M^t\ O^{t-1})\\ \times P(V^t_{(x,y)} \mid O^t_{(x,y)})\end{array}\right]\\ \times\prod_{i=1}^{N} P(T^t_i \mid M^t\ O^t\ T^{t-1}_i)\end{array}\right]
\end{aligned}
$$

The additional factors $P(T^0_i)$ are priors over the positions of the targets that can be set according to the starting position of the targets as shown in the cueing phase.

The dynamic model of targets is very similar to the dynamic model of objects but with the occupancy grid on objects as observation (see Appendix B.2 for details).

At each time step, the relevant state of the representation can be summarized by the following question for each target $i \in [\![1, N]\!]$ at each timestep $t \in [\![1, t_{\max}]\!]$:

$$P(T^t_i \mid V^{1\to t}\ M^{1\to t}) \tag{2}$$

Bayesian inference leads to the following expression for this question:

$$P(T_i^t \mid V^{1\to t}\ M^{1\to t}) \propto \sum_{T_i^{t-1}} \left[\begin{array}{l} \sum_{O^t} \left[\begin{array}{l} P(T_i^t \mid M^t\ O^t\ T_i^{t-1}) \\ \times P(O^t \mid V^{1\to t}\ M^{1\to t}) \end{array} \right] \\ \times P(T_i^{t-1} \mid V^{1\to t-1}\ M^{1\to t-1}) \end{array} \right]$$

where $P(T_i^{t-1} \mid V^{1\to t-1}\ M^{1\to t-1})$ is the result of the same inference at the preceding timestep, $P(O^t \mid V^{1\to t}\ M^{1\to t})$ the result of question 1 at the same timestep. The summation of the whole grid, which is still computationally intensive, can be approximated as above, by separating the cells.

Both questions 1 and 2 are the current knowledge about the visual scene that can be inferred from the past observations and movements, and the hypotheses of our model.

2.2 Decision models

Based on this knowledge, the observer has to decide where to look next in order to solve the task. We propose different models in order to test different hypotheses. First, we make the hypothesis that this representation model is useful for producing eye movements. To test this hypothesis, we compare a model that does not use the representation with one that does.

Then, the main hypothesis is that uncertainty, explicitly taken into account, can help in the decision of eye movement. Therefore, we compare a model that does not take into account explicitly the uncertainty with one that does.

In the end, we need to specify three models: one that does not use the representation model (π_A), one that uses the representation model without explicitly taking into account uncertainty (π_B), and finally one that uses the representation model and explicitly takes into account uncertainty (π_C). Each model π_k will infer a probability distribution on the next eye movement represented by a new variable $C^t \in \mathcal{G}$ at each time $t \in [\![1, t_{\max}]\!]$:

$$P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_k)$$

This variable is the model's homologue to the motor cells found in LIP, FEF and SC.

2.2.1 Constant model

This model is a baseline for the other models. We look for the best static probabilistic distribution that can account for the experimental eye movement. Formally it is specified as being independent on time and on the observations:

$$\forall t \in [\![1, t_{\max}]\!], \quad P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_A) = P(C^t \mid \pi_A) = P(C^1 \mid \pi_A)$$

In these conditions, it can be shown that the best distribution $P(C^1 \mid \pi_A)$, according to the measure defined Sect. 3.1, assigns the probability of each individual discretized motion to be equal to its frequency in the experimental data.[2] Therefore, we learned this distribution from our experimental data, using only a randomly selected subset in order not to overfit our models.

2.2.2 Targets positions

The second model we propose uses the knowledge from the representation layer to determine its eye movements. More precisely, it tends to look at locations where targets are close to another, in a kind of fusion process. Its prior will follow the statistical distribution of eye movements and the likelihood will be based on the distributions on the targets location inferred in the representation layer.

The decomposition is as follows:

$$P(CVMT \mid \pi_B) = \prod_{t=1}^{t_{\max}} \left[\begin{array}{l} P(V^t\ M^t \mid \pi_B) \\ \times \prod_{i=1}^{N} P(T_i^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \\ \times P(C^t \mid T^t\ \pi_B) \end{array} \right]$$

where:

- $P(V^t\ M^t \mid \pi_B)$ is an arbitrary prior that is not used in the inference,
- $P(T_i^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B)$ is the result of inference 2,
- $P(C^t \mid T^t\ \pi_B)$ is the result of the inference in a fusion submodel over the targets that yields:

$$P(C^t \mid T^t\ \pi_B) \propto P(C^t \mid \pi_A) \prod_{i=1}^{N} P(T_i^t \mid C^t)$$

where $P(C^t \mid \pi_A)$ is the prior taken from the constant model and $P(T_i^t \mid C^t)$ a distribution centered on C^t that expresses a proximity between C^t and T_i^t (concretely a Gaussian distribution centered on C^t).

With this model, the distribution on eye movement can be computed with the following expression:

$$P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \propto P(C^t \mid \pi_A) \times \prod_{i=1}^{N} \sum_{T_i^t} \left[\begin{array}{l} P(T_i^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \\ \times P(T_i^t \mid C^t) \end{array} \right]$$

[2] When restricted to time independence and assuming a uniform prior over such models, our measure is a multinomial likelihood which leads to a Dirichlet distribution according to the experimental frequencies. The maximum of this Dirichlet distribution is the histogram of the experimental frequencies.

In short, this model is the product between the prior on eye movement and each distribution on the targets convolved by a Gaussian distribution. This expression shows that this model is attracted towards the targets but without necessarily looking at one in particular as balance between the distributions on the targets can lead to a peak in some weighted sum of their locations.

2.2.3 Uncertainty model

The behavior of the preceding model is influenced by uncertainty insofar as the incentive to look near a given target is higher for a more certain location of this target. As for any Bayesian model, uncertainty is handled as part of the inference mechanism: as a mean to describe knowledge.

In this third model, we propose to include uncertainty as a variable to reason about: as the knowledge to be described. The rationale is simply that it is more efficient to gather information when and where it lacks than when and where there is less uncertainty.

Therefore, we introduce a new set of variables $I^t_{(x,y)} \in [0, 1]$, representing an index of the uncertainty at cell $(x, y) \in \mathcal{G}$ at time $t \in [\![1, t_{\max}]\!]$. Any index can fit as long as we can correlate the value of this uncertainty index with the actual uncertainty.

For simplification, we choose our uncertainty indices to be equal to this probability of occupancy, as we represent occupancy as binary variables. The relation between this uncertainty index (probability distribution) and uncertainty is such that a probability near $\frac{1}{2}$ represents a high uncertainty whereas a probability near 0 or 1 represent a low uncertainty. Other spaces can be chosen for these variables, such as entropy, but we keep the probability distribution to simplify our computations.

As mentioned above, this model is structured around a prior probability of motion which is filtered by these uncertainty variables in order to enhance the probability of eye movements towards uncertain regions. The prior probability is the result of the preceding model π_B.[3]

The decomposition of this model is as follows:

$$P(CVMI \mid \pi_C) = \prod_{t=1}^{t_{\max}} \left[\begin{array}{l} P(V^t\ M^t \mid \pi_C) \\ \times P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \\ \times \prod_{(x,y)\in\mathcal{G}} P(I^t_{(x,y)} \mid C^t\ \pi_C) \end{array} \right]$$

where:

- $P(V^t\ M^t \mid \pi_C)$ is an arbitrary prior that is not used in the inference,
- $P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B)$ is the result of the previous model,
- $P(I^t_{(x,y)} \mid C^t\ \pi_C)$ is a beta distribution that expresses that for a given eye movement proposal C^t, $I^t_{C^t}$ is more likely near $\frac{1}{2}$ and distribution on $I^t_{(x,y)}$ for $(x, y) \neq C^t$ is uniform.

This model computes the posterior probability distribution on next eye movement using the following expression:

$$P(C^t \mid V^{1\to t}\ M^{1\to t}\ I^{1\to t}\ \pi_C) \propto P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \times P(I^t_{C^t} \mid C^t\ \pi_C)$$

where:

$$\forall (x, y), t \in \mathcal{G} \times [\![1, t_{\max}]\!],\quad I^t_{(x,y)} = P(O^t_{(x,y)} \mid V^{1\to t}\ M^{1\to t})$$

as computed by Eq. 1.

This model filters the eye movement distribution computed by the second model, in order to enhance the probability distribution in the locations of high uncertainty.

3 Results

The output of our models is a probability distribution over the eye position at each timestep. For such complex objects, there are neither significance test nor an appropriate sensitivity analysis and the comparison is done using their respective likelihood. However the likelihood is highly dependent on the size of the data set. Therefore we first introduce a comparison method that does not depend on the size of the data set. Then we present their results and comment them with respect to the specific behavior of each model. Finally, we illustrate the main differences between the various models by giving examples of specific situations.

3.1 Comparison method

The decision models compute a probability distribution over the possible eye movements at one moment, based on past observations and their respective hypotheses (Fig. 4). We can therefore compute, for each model, the probability of the actual eye movements recorded from subjects in a given situation, as well as the probability of the whole set of recordings with an additional independency assumption.

Probability values are only relative measures as, when the possibilities are numerous, they tend to be very small. However, their comparison across models (which share the same number of possibilities) indicates which model is a better predictor of the recorded eye movements. This process is known as the *Maximum Likelihood* method.

[3] This is a matter of presentation of the model. The complete expression of π_C can be written without reference to model π_B but the addition of uncertainty would be less clear.

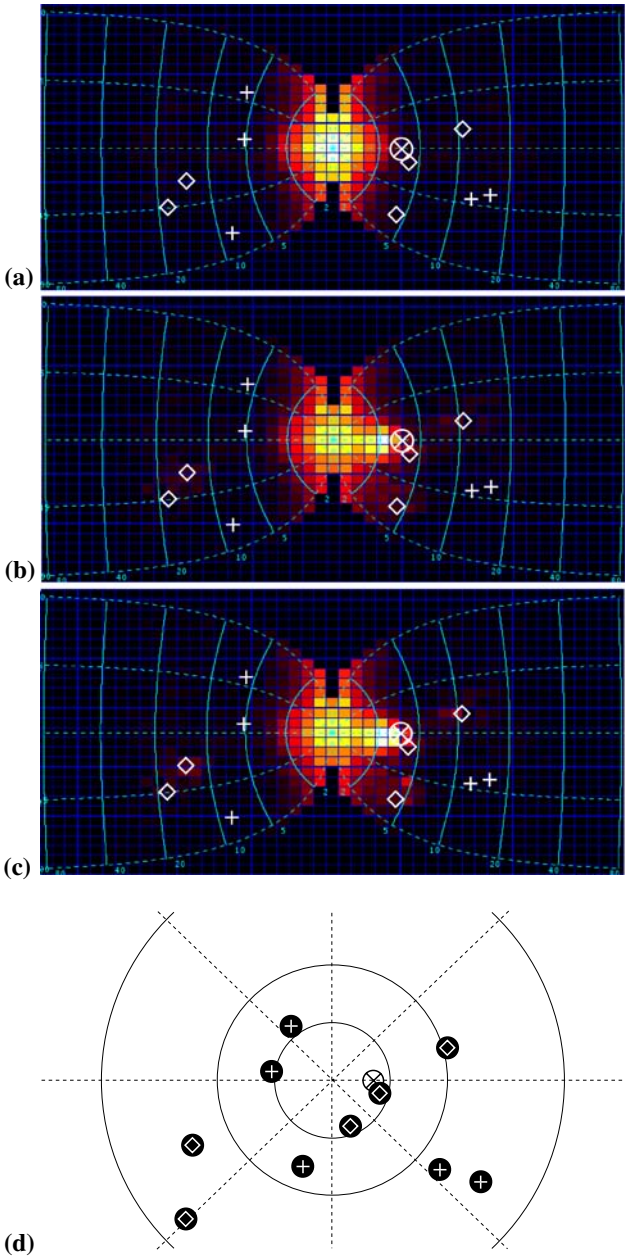

Fig. 4 Example of probability distributions computed by each decision model in the same configuration. The two halves of the representations are drawn side-by-side. The *plain lines* are the iso-eccentricities and the *dotted lines* are the iso-directions. The *brightness* of the cell indicates the probability of the associated eye movement: a dark cell for a low probability and a white cell for a high probability for the eye movement toward this cell. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** is the probability distribution of constant model. **b** shows the probability distribution for the target model that shows a preference for the targets. **c** shows the probability distribution for the uncertainty model that highlights some of the targets. **d** shows the position of the targets and distractors in the visual field. Note that the probability distributions for model **c** favors the next eye movement

However, except in very special cases, the likelihood of a model would decrease exponentially toward zero with the increase of the number of trials, while the likelihood ratio between two models will diverge or converge exponentially toward zero. Therefore, we compare our decision models using the geometric mean of the likelihood of the observed eye movements over each trial. The geometric mean allows to be a substitute for the complete likelihood, as it is its Nth root where N is the total number of trials, while providing a measure converging to a non-zero value as the number of trials grows.

More precisely, let c_n^t be the tth eye movement recorded during trial n. The likelihood of a model π for trial n is:

$$\prod_{t=1}^{t_{\max}} P([C^t = c_n^{t+1}] \mid v_n^{1\to t}\ c_n^{1\to t}\ \pi)$$

The global likelihood of model π is:

$$\prod_{n=1}^{N}\prod_{t=1}^{t_{\max}} P([C^t = c_n^{t+1}] \mid v_n^{1\to t}\ c_n^{1\to t}\ \pi)$$

Finally we define our measure μ to be the geometric mean of the likelihood over all the trials:

$$\mu(\pi) = \sqrt[N]{\prod_{n=1}^{N}\prod_{t=1}^{t_{\max}} P([C^t = c_n^{t+1}] \mid v_n^{1\to t}\ c_n^{1\to t}\ \pi)} \tag{3}$$

3.2 Results and analysis

The data set is gathered from 11 subjects with 110 trials each for a total of 1,210 trials (Tanner et al. 2007). Each trial was regularly discretized in time in $t_{\max} = 24$ observations (with a timestep of 200 ms) for a grand total of 29,040 data points. The eye movement variable M^t is build from the difference in gaze position between two successive timesteps. Part of the data set (124 random trials) was used to determine the parameters of the various models and the results are computed on the remaining $N = 1{,}089$ trials.

Table 1 presents the ratio of the measure for each pair of our three decision models computed for this data set. It shows that the model which generates motion with the empiric probability distribution but without the representation layer is far less probable than the other two (by respectively a factor 280 and 320). This shows that, as expected, the representation layer is useful in deciding the next eye movement.

Table 1 Ratio of the measures for each pair of models

Model	Model		
	Constant (π_A)	Target (π_B)	Uncertainty (π_C)
Constant (π_A)	1	280	320
Target (π_B)	3.5×10^{-3}	1	1.14
Uncertainty (π_C)	3.1×10^{-3}	0.87	1

Table 1 further shows that the model taking explicitly into account uncertainty is 14% more likely than the model that does not. This is in favor of our hypothesis that taking explicitly into account uncertainty is helpful in deciding the next eye movement.

As explained above, the choice of the geometric mean prevents the measure to converge toward zero and prevents their ratios to raise exponentially as the number of trials grows. In our case, the likelihood ratio between the model with explicit uncertainty and the one without is 4.9×10^{63}. With half the trials, this likelihood ratio is the square root, that is only 7.0×10^{31}. This shows that the likelihood ratio is indeed not a stable measure with respect to the number of trials. We preferred a stable measure in order to have a more meaningful value.

3.3 Typical situations

These results show a global agreement of the model with the actual eye movements of the human participants. However, there are some configurations where the models can have different relative performances. The analysis of such examples can shed some light on the behavior of the various decision models we proposed.

3.3.1 Examples where π_C is more likely than π_B

The global result shows that it is better to take into account uncertainty explicitly for the choice of the eye movement. We can further investigate by looking at the frames where the difference in the likelihood is greatest.

We isolated two different categories of configurations where model π_C was especially better than model π_B, exemplified in Fig. 5. The first category consists in scenes where a target and a distractor are in a close vicinity and the eye movement of the participant is around those objects (Fig. 5a). In these case, the target model is simply attracted by the target whereas the uncertainty model is additionally attracted by both objects due to their uncertainty.

The second category consists in occurrences of an eye movement towards a distractor (see Fig. 5b). In this case, the target model has no incentive for looking at this location whereas there is always some uncertainty to investigate for model π_C.

3.3.2 Examples where π_B is more likely than π_C

Even if the global results are in favor of the model with explicit uncertainty, there are cases where the target model better predicts the eye movements. This happen mainly when the eye movements occur in the middle of several targets but not on a particular one (Fig. 6a). In this case, the fusion on the targets employed by model π_B can present a maximum

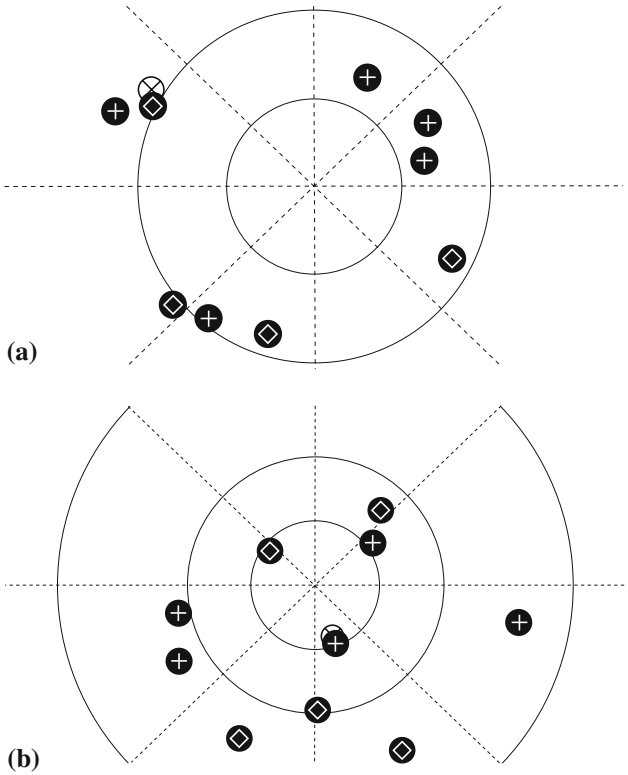

Fig. 5 Examples of eye movements better predicted by model π_C than model π_B. The scene is presented in an eye centered reference frame. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** The actual eye movement occurs towards both a target and a distractor. **b** The actual eye movement occurs towards an isolated distractor

in a center of mass of the targets, whereas the absence of objects—and therefore the low uncertainty—will lower the probability of this particular eye movement by model π_C.

Figure 6b illustrates a second interesting case. The eye movement occurs in between a target and a distractor. However, the occupancy grid at that time (Fig. 6c) shows that the target is moving and the eye movement is near the previous position of the target shown by a peak of occupancy in the corresponding cell. Therefore the eye movement is near the representation of the target. On the other hand, there is also a great patch near the center of the visual field with a moderate level of uncertainty where, consequently, model π_C predicts a high probability of eye movement.

3.3.3 Examples where π_A is more likely than π_B or π_C

Finally, the constant model can also be the most likely one for some particular configurations and movements. This occurs mostly for fixations that are not directed to objects (for example Fig. 7a). Indeed model π_A is simply the global distribution of eye movements that are mostly of low amplitude (see Fig. 4a) and the other models are mostly attracted to targets or the uncertainty attached to objects.

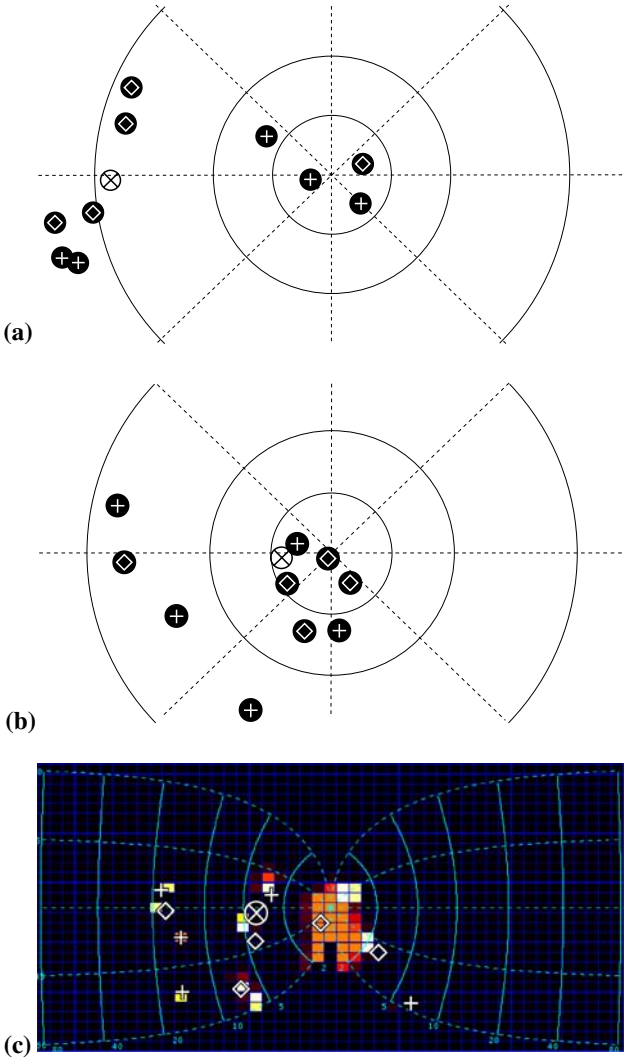

Fig. 6 Examples of eye movements better predicted by model π_B than model π_C. The scene is presented in an eye centered reference frame. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** The actual eye movement occurs in between several targets. **b** The actual eye movement occurs towards an isolated distractor. **c** Occupancy grid for the same configuration depicted in **b** showing the eye movement is near the past location of the target

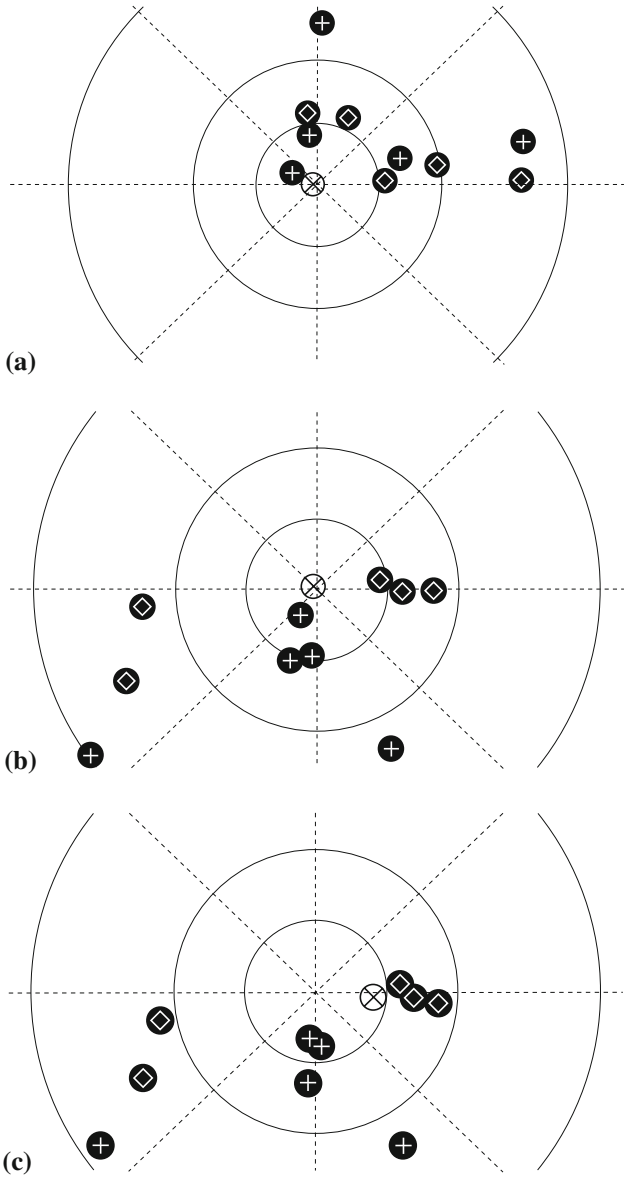

Fig. 7 Examples of eye movements better predicted by model π_A than models π_B or π_C. The scene is presented in an eye centered reference frame. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** The actual eye movement is a fixation without object. **b** The actual eye movement is also a fixation although there is a group of targets on the right. **c** Situation following **b** where the eye movement is towards the group of targets

Figure 7b shows another occurrence of this situation with a group of target on the right towards which the other models predict a high probability of movement. It happens that, on the next frame, shown Fig. 7c, for which the situation is similar, the participant looked towards this group of targets, as predicted by both models π_B and π_C.

4 Conclusion and discussion

As a conclusion, we propose a Bayesian model with two parts: a representation of the visual scene, and a decision model based on the state of the representation. The representation both tracks the occupancy of the visual scene as well as the locations of the targets. Based on this representation, we tested several decision models and we have shown that the model that takes explicitly into account the uncertainty better fitted the eye movements recorded from subjects participating a psychophysics experiment.

In addition, the eye movement frequency shows that, most of the times, the eye movements are of low amplitude, indicating either fixation or slow pursuit of an object. In these cases, the constant model has a likelihood comparable with or even

sometimes greater than the other two. Thus the difference is due to the saccadic events, when the target and uncertainty model have a higher likelihood than the constant one which assigns a lower probability as the eccentricity grows. On the other hand, the difference between the target model and the uncertainty model is due to the filtering of the eye movements distribution from the target model by the uncertainty. The difference is less substantial than for the constant model as the uncertainty associated to the targets are often similar (isolated targets with comparable movement profiles). It could be interesting to enrich the stimuli in order to manipulate uncertainty more precisely.

The stimulus is adapted from the classical MOT task used primarily to study attention. Our model uses a set of variables to track the position of the targets. This set of variable is fixed and finite (five in our model), which means our model can only track as much targets as its number of target position variables. The human subjects, however, are also informed about the number of targets in the instructions. Experimental evidence suggests that human performance drops if the number of target gets too high. For the particular experimental design we used, the maximum number of targets consistently tracked was 5, which justifies our choice of the number of target variables. Other experimental studies suggest that this maximum number of target is not fixed and seems to depend on factors such as speed and spacing of the objects (Alvarez and Franconeri 2007). In addition, each of our target variables cover the whole visual field (encoded in the logcomplex mapping) although there are works indicating that some representation capacities are separated across the hemifields (Alvarez and Cavanagh 2005). It could be interesting to test this in our model with a set of target variables for the left part and another for the right part. However, due both to eye movements and targets movements, the targets sometimes change side, implying some additional mechanism of communication between these variables.

Finally, one of the main features of our model is to place all computations and representation in the logcomplex mapping found in the neurophysiology of some retinotopic maps. To our surprise, we found in the psychophysical data that the distribution of the objects positions is quite uniform in the logcomplex mapping. This suggests a particular strategy for the eye movements. One interpretation could be that the eye movements are chosen in order to maximize the use of the representation: that is, so that the objects are uniformly distributed in this representation. This seems to be an indirect confirmation that eye movements are governed by structures using this particular mapping.

Acknowledgments The authors acknowledge the support of the European Project BACS (Bayesian Approach to Cognitive Systems), FP6-IST-027140. The authors thank Luiz Canto-Pereira, Heinrich Bülthoff, and Cristóbal Curio for their involvement in the experimental aspects of this work. The authors also thank warmly Julien Diard for the insightful discussions about the preliminary model design.

Appendix A: Experimental protocol

This experiment is an adaptation of the classical MOT paradigm from Pylyshyn and Storm (1988) (see Fig. 3) but with eye movements. In the original task, participants were asked to keep track of a given number of targets among identical distractors as they all move independently on the screen. Participants had to keep their gaze at a fixating point located on the center of the screen. Therefore the targets will occasionally be located in the periphery of the visual field, in the low resolution areas of the visual field. Therefore we expect eye movements to occur in order to keep track of targets.

A.1 Materials and methods

A.1.1 Participants

Eleven subjects participated in the experiment with normal or corrected vision. Each session consists of 110 trials.

A.1.2 Apparatus

The stimulus is presented on a calibrated 21″ Sony CPD-500 CRT monitor with a refresh rate of 100 Hz and a resolution of $1{,}024 \times 768$. Participants are positioned in front of the monitor at a distance of 65 cm; at this distance the display subtended a visual angle of 33° by 25°. A chin rest ensures that no head movement occurs during the experimental session. All experimental sessions are performed in a sound attenuated room with controlled artificial lighting. Eye movements are recorded by an eye tracker system (EyeLink II, SR Research Ltd.) with a sampling rate of 250 Hz and an accuracy of ca. 0.3°. The model was simulated offline with a timestep of 200 ms using the difference in eye position between two timesteps. No analysis of saccades, micro-saccades, pursuit or fixation was needed in this respect.

A.2. Procedure

The display consists of ten identical objects, each one a white circle subtending 1° of visual angle, with a luminance of 90 cd/m^2 against a black background, in a room illuminated with diffuse D65 light (70 cd/m^2).

Targets and distractors are identical with the exception of the initial phase in the beginning of each trial. In this phase, five targets are cued by a series of three flashes, with a total duration of 1,080 ms. After this initial phase, all objects begin

to move in different directions, chosen from among 8 directions of the compass with a mean velocity of 5.1° per second.

The objects have random initial locations, directions and speeds during trials but are constrained to keep a minimum distance of 1.5° (Pylyshyn and Storm 1988).

Trials last 5 s and on the end of each trial participants are asked to select targets with a mouse.

More details can be found in the description of experiment B in (Tanner et al. 2007, paper in preparation).

Appendix B: Dynamic models

B.1 Dynamic object model

This dynamic model provides the transition probability distribution $P(O^t_{(x,y)} \mid M^t\ O^{t-1})$ that governs the evolution of the grid with the remapping capability. In order to stress the issue of the logcomplex mapping, we explicitly refer to the visual coordinates (ρ, θ) as well as the logcomplex coordinates (x, y). We also consider coordinates $(\rho, \theta)_{\text{ant}}$ and $(x, y)_{\text{ant}}$ to denote coordinates at the previous time step. In the end, the decomposition is as follows:

$$\begin{aligned}
&P((x,y)\ (x,y)_{\text{ant}}\ (\rho,\theta)\ (\rho,\theta)_{\text{ant}}\ O^t_{(x,y)}\ O^{t-1}\ M^t)\\
&\quad = P((x,y))P(M^t)P(O^t_{(x,y)})P((\rho,\theta) \mid (x,y))\\
&\qquad \times P((\rho,\theta)_{\text{ant}} \mid (\rho,\theta)\ M^t)P((x,y)_{\text{ant}} \mid (\rho,\theta)_{\text{ant}})\\
&\qquad \times \prod_{(x',y')} P(O^{t-1}_{(x',y')} \mid O^t_{(x,y)}\ (x,y)_{\text{ant}})
\end{aligned}$$

where:

- $P((x, y))$ is an arbitrary unused distribution;
- $P(M^t)$ is an arbitrary unused distribution;
- $P(O^t_{(x,y)})$ is a uniform distribution;
- $P((\rho, \theta) \mid (x, y))$ is a uniform distribution on the inverse image of the position (x, y) by the logcomplex mapping;
- $P((\rho, \theta)_{\text{ant}} \mid (\rho, \theta)\ M^t)$ is a Dirac distribution on the image of (ρ, θ) by eye movement M^t;
- $P((x, y)_{\text{ant}} \mid (\rho, \theta)_{\text{ant}})$ is a Dirac distribution on the cell corresponding to position $(\rho, \theta)_{\text{ant}}$;
- $P(O^{t-1}_{(x',y')} \mid O^t_{(x,y)}\ (x^{-1}, y^{-1}))$ is a transition matrix that states there is a great probability to keep the same occupancy if $(x', y') = (x, y)_{\text{ant}}$, and is a uniform distribution otherwise.

This model is used to compute the question $P(O^t_{(x,y)} \mid M^t\ O^{t-1})$ using the following expression:

$$\begin{aligned}
&P(O^t_{(x,y)} \mid O^{t-1}\ M^t)\\
&\quad \propto \sum_{(\rho,\theta)} P((\rho,\theta) \mid (x,y))P(O^{t-1}_{(\hat{x},\hat{y})} \mid O^t_{(x,y)}\ (\hat{x},\hat{y}))
\end{aligned}$$

where $(\hat{x}, \hat{y})$ are the coordinates of the cell corresponding to the image of (ρ, θ) by eye motion M^t.

This summation can be implemented by sampling the distribution $P((\rho, \theta) \mid (x, y))$.

B.2 Dynamic target model

This dynamic target model is common to every target and combines both the prediction of the position of the target based only on eye movement (remapping) and the update of this position according to the occupancy grid. It provides the distribution $P(T^t_i \mid T^{t-1}_i\ O^t\ M^t)$ used in the representation model.

The decomposition is as follows:

$$\begin{aligned}
&P(T^t_i\ T^{t-1}_i\ (\rho,\theta)\ (\rho,\theta)_{\text{ant}}\ M^t\ O^t)\\
&\quad = P(T^t_i)P(M^t)P((\rho,\theta) \mid T^t_i)\\
&\qquad \times P((\rho,\theta)_{\text{ant}} \mid (\rho,\theta)\ M^t)P(T^{t-1}_i \mid (\rho,\theta)_{\text{ant}})\\
&\qquad \times \prod_{(x,y)} P(O^t_{(x,y)} \mid T^t_i)
\end{aligned}$$

where:

- $P(T^t_i)$ is a uniform distribution;
- $P(M^t)$: is an arbitrary unused distribution;
- $P((\rho, \theta) \mid T^t_i)$ is a uniform distribution on the inverse image of the position T^t_i by the logcomplex mapping;
- $P((\rho^{-1}, \theta^{-1}) \mid (\rho, \theta)\ M^t)$ is Dirac distribution on the image of (ρ, θ) by eye movement M^t;
- $P(T^{t-1}_i \mid (\rho, \theta)_{\text{ant}})$ is a Dirac on the cell corresponding to position $(\rho, \theta)_{\text{ant}}$;
- $P(O^t_{(x,y)} \mid T^t_i)$ states that it is more probable to have an occupied cell in a neighborhood of T^t_i, and that it is uniform elsewhere.

This model is used to compute the question $P(T^t_i \mid T^{t-1}_i\ O^t\ M^t)$ with the following expression:

$$\begin{aligned}
&P(T^t_i \mid T^{t-1}_i\ M^t\ O^t)\\
&\quad \propto \left|\mathcal{E}(T^{t-1}_i, M^t)\right| \prod_{(x,y)} P(O^t_{(x,y)} \mid T^t_i)
\end{aligned}$$

where $\left|\mathcal{E}(T^{t-1}_i, M^t)\right|$ is the size of the set of the polar positions (ρ, θ) that are in relation with T^{t-1}_i by the eye movement M^t. This set can be obtained by sampling like in the dynamic model.

Appendix C: Implementation details

The models presented are implemented in the Java language. In all the examples, the grid $\mathcal{G}$ is composed of 24 × 29 cells

for each hemifield and we used a timestep of 200 ms for the representation and decision models.

Additionally, some of the probability distributions described as factors in the decompositions are parametric forms that need precise values to be involved in actual computations. We explored the parametrical space and evaluated each parameter set with our measure computed on a subset of the experimental data.

Finally, in the representation model, the observation model $P(V^t_{(x,y)} \mid O^t_{(x,y)})$ is a 2×2 matrix with value 0.9 on the diagonal and 0.1 elsewhere

$$\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}.$$

The transition matrix of the dynamic model is

$$\begin{pmatrix} 0.95 & 0.1 \\ 0.05 & 0.9 \end{pmatrix}.$$

The target observation model $P(O^t_{(x,y)} \mid T^t_i)$ is of the form $0.5 + \frac{0.25}{1+\left(\frac{d((x,y),T^t_i)}{0.02}\right)^2}$ for an occupied cell and $0.5 - \frac{0.25}{1+\left(\frac{d((x,y),T^t_i)}{0.02}\right)^2}$ otherwise with $d((x, y), T^t_i)$ the distance between cell (x, y) and position T^t_i in mm. The target fusion model $P(T^t_i \mid C^t)$ is a mixture between a Gaussian and a uniform distribution: $\propto 0.25 + \exp -\frac{d(T^t_i,C^t)^2}{0.25}$. In the uncertainty decision model, the uncertainty fusion distribution $P(I^t_{(x,y)} \mid C^t\, \pi_C)$ is a symmetrical beta distribution with parameter 0.075.

References

Alvarez GA, Cavanagh P (2005) Independent resources for attentional tracking in the left and right visual hemifields. Psychol Sci 16(8):637–643

Alvarez GA, Franconeri SL (2007) How many objects can you attentively track? Evidence for a resource-limited tracking mechanism. J Vis 7(13):1–10. http://journalofvision.org/7/13/14/

Barash S, Bracewell R, Fogassi L, Gnadt J, Andersen R (1991a) Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. J Neurophysiol 66(3): 1095–1108

Barash S, Bracewell R, Fogassi L, Gnadt J, Andersen R (1991b) Saccade-related activity in the lateral intraparietal area. II. Spatial properties. J Neurophysiol 66(3):1109–1124

Ben Hamed S, Duhamel JR, Bremmer F, Graf W (2001) Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. Exp Brain Res 140:127–144

Bessière P, Laugier C, Siegwart R (2008) Probabilistic reasoning and decision making in sensory-motor systems. Springer, Berlin

Bozis A, Moschovakis A (1998) Neural network simulations of the primate oculomotor system III. An one-dimensional, one-directional model of the superior colliculus. Biol Cybern 79:215–230

Cavanagh P, Alvarez GA (2005) Tracking multiple targets with multifocal attention. Trends Cogn Sci 9(7): 349–354

Droulez J, Berthoz A (1991) A neural network model of sensoritopic maps with predictive short-term memory properties. Proc Natl Acad Sci 88:9653–9657

Elfes A (1989) Occupancy grids: a probabilistic framework for robot perception and navigation. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, USA

Fehd HM, Seiffert AE (2008) Eye movements during multiple object tracking: where do participants look. Cognition 108(1):201–209

Gnadt J, Andersen R (1988) Memory related motor planning activity in the posterior arietal cortex of the macaque. Exp Brain Res 70(1):216–220

Goldberg M, Bruce C (1990) Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal. J Neurophysiol 64(2):489–508

Herrero L, Rodríguez F, Salas C, Torres B (1998) Tail and eye movememnts evoked by electrical microstimulation of the optic tectum in goldfish. Exp Brain Res 120:291–305

Krauzlis R (2004) Recasting the smooth pursuit eye movement system. J Neurophysiol 91(2):591–603

Lebeltel O, Bessière P, Diard J, Mazer E (2004) Bayesian robots programming. Auton Robots 16(1):49–79

Mays L, Sparks D (1980) Dissociation of visual and saccade-related responses in superior colliculus neurons. J Neurophysiol 43(1):207–232

McIlwain J (1976) Large receptive fields and spatial transformations in the visual system. In: Porter R (ed) Neurophysiology II, Int Rev Physiol, vol 10. University Park Press, Baltimore, pp 223–248

McIlwain J (1983) Representation of the visual streak in visuotopic maps of the cat's superior colliculus: influence of the mapping variable. Vis Res 23(5):507–516

Mitchell J, Zipser D (2003) Sequential memory-guided saccades and target selection: a neural model of the frontal eye fields. Vis Res 43:2669–2695

Moschovakis A, Scudder C, Highstein S (1996) The microscopic anatomy and physiology of the mammalian saccadic system. Prog Neurobiol 50:133–254

Ottes F, van Gisbergen JA, Eggermont J (1986) Visuomotor fields of the superior colliculus: a quantitative model. Vis Res 26(6): 857–873

Pylyshyn Z, Storm R (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism. Spatial Vis 3(3):1–19

Robinson D (1972) Eye movements evoked by collicular stimulation in the alert monkey. Vis Res 12:1795–1808

Schwarz E (1980) Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. Vis Res 20:645–669

Scudder C, Kaneko C, Fuchs A (2002) The brainstem burst generator for saccadic eye movements. A modern synthesis. Exp Brain Res 142:439–462

Siminoff R, Schwassmann H, Kruger L (1966) An electrophysiological study of the visual projection to the superior colliculus of the rat. J Comp Neurol 127:435–444

Sommer M, Wurtz R (2000) Composition and topographic organization of signals sent from the frontal eye fields to the superior colliculus. J Neurophysiol 83:1979–2001

Tanner T, Canto-Pereira L, Bülthoff H (2007) Free vs. constrained gaze in a multiple-object-tracking-paradigm. In: 30th European Conference on Visual Perception, Arezzo, Italy

Wurtz R, Sommer M, Paré M, Ferraina S (2001) Signal transformation from cerebral cortex to superior colliculus for the generation of saccades. Vis Res 41:3399–3412

Zelinsky GJ, Neider MB (2008) An eye movement analysis of multiple object tracking in a realistic environment. Vis Cogn 16(5):553–566

6.4 (Dollé et al, 2010)

Article accepté le 21 juin 2010 pour publication dans Biological Cybernetics.

Biol Cybern
DOI 10.1007/s00422-010-0400-z

ORIGINAL PAPER

Biological Cybernetics

Path planning versus cue responding: a bio-inspired model of switching between navigation strategies

Laurent Dollé · Denis Sheynikhovich · Benoît Girard · Ricardo Chavarriaga · Agnès Guillot

Received: 8 January 2010 / Accepted: 21 June 2010

Abstract In this article, we describe a new computational model of switching between path-planning and cue-guided navigation strategies. It is based on three main assumptions: (i) the strategies are mediated by separate memory systems that learn independently and in parallel; (ii) the learning algorithms are different in the two memory systems—the cue-guided strategy uses a temporal-difference (TD) learning rule to approach a visible goal, whereas the path-planning strategy relies on a place-cell-based graph-search algorithm to learn the location of a hidden goal; (iii) a strategy selection mechanism uses TD-learning rule to choose the most successful strategy based on past experience. We propose a novel criterion for strategy selection based on the directions of goal-oriented movements suggested by the different strategies. We show that the selection criterion based on this "common currency" is capable of choosing the best among TD-learning and planning strategies and can be used to solve navigational tasks in continuous state and action spaces. The model has been successfully applied to reproduce rat behavior in two water-maze tasks in which the two strategies were shown to interact. The model was used to analyze competitive and cooperative interactions between different strategies during these tasks as well as relative influence of different types of sensory cues.

Keywords Computational model · Spatial navigation · Strategy switch · Parallel memory systems · Action selection

Laurent Dollé, Denis Sheynikhovich—First authorship shared.

L. Dollé (✉) · D. Sheynikhovich · B. Girard · A. Guillot
Institut des Systèmes Intelligents et de Robotique, UPMC CNRS UMR 7222, 4 Place Jussieu, 75252 Paris Cedex 05, France
e-mail: laurent.dolle@upmc.fr

Present Address:
D. Sheynikhovich
Neurobiology of Adaptive Processes UPMC, CNRS UMR 7102,
9 quai St. Bernard, 75005 Paris, France

R. Chavarriaga
Defitech chair on Non-Invasive Brain-Computer Interface (CNBI), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

1 Introduction

An increasing number of behavioral research studies focus on the capacity of animals to switch between different navigation strategies when it is required by the environmental circumstances (see Franz and Mallot 2000; White 2004; Arleo and Rondi-Reig 2007; Khamassi 2007, for reviews). The majority of these articles explore the interactions between response- and place-based strategies (Packard and McGaugh 1996; Devan and White 1999; Roberts and Pearce 1999; Gibson and Shettleworth 2005; Rich and Shapiro 2009). Response-based strategies are thought to learn associations between sensory cues and actions linked with reward, whereas place-based strategies use a form of spatial representation to store the goal position and plan a path to it. Experimental evidence in support of such a separation between navigational strategies comes from lesion's studies that gave rise to the theory of parallel memory systems in the brain of the rat (Packard et al. 1989; McDonald and White 1993; Devan and White 1999; Kim and Baxter 2001; White and McDonald 2002; White 2004; Burgess 2008). According to this theory, the dorsolateral striatum (DLS) is involved in the control of response-based strategies by means of a slow and inflexible "trial and error" learning, whereas place-based strategies are mediated by the hippocampus (Hc) and other neural structures to which it projects, such as prefrontal

Published online: 09 July 2010

cortex (PFC) (Mizumori 2008; Jankowski et al. 2009; White 2009). Learning in the Hc-dependent pathway is considered to be rapid and flexible (Granon and Poucet 1995; Yin and Knowlton 2004; Grahn et al. 2008).

The existence of two (or more) parallel memory systems mediating different behavioral strategies raises a question of when one or other strategy takes control over behavior. Experimental evidence suggests that different memory systems favor separate sets of sensory cues: DLS-mediated system mostly uses proximal cues (e.g., visible platform in the Morris Water Maze, or intra-maze landmark signaling the platform position), whereas Hc-mediated system encodes configurations of distal cues (like extra-maze landmarks and environmental boundaries) (McDonald et al. 2004; Hartley and Burgess 2005; Doeller and Burgess 2008; Doeller et al. 2008; Leising and Blaisdell 2009; Blaisdell 2009; Pearce 2009). Distal cues and environmental boundaries can be used to form a spatial representation encoded in the activities of location selective neurons (termed "Place Cells") residing in the Hc (O'Keefe and Dostrovsky 1971; O'Keefe and Nadel 1978; Redish 1999; Save and Poucet 2000; Kelly and Gibson 2007). The question raised by these studies is how different types of sensory cues influence ongoing behavior, including strategy selection.

Interactions between multiple navigation strategies when two or more of them can be used at the same time is often analyzed in terms of competition and cooperation. *Competition* between two memory systems (and hence, the corresponding strategies) is demonstrated when a lesion of one of the systems entails an improvement of the learning of the other, while *cooperation* implies that such a lesion leads to the impairment of the other system's performance (Kim and Baxter 2001; Gold 2004). In the spatial domain, competition or cooperation between navigational strategies are respectively observed when one of the strategies perturbs (Packard and McGaugh 1992; Pearce et al. 1998; Chang and Gold 2003; Canal et al. 2005) or facilitates (McDonald and White 1994; Hamilton et al. 2004; Voermans et al. 2004) the other one for reaching the goal. The analysis of switching between place- and response-based strategies suggests that they can interact both across and within experimental trials (Pearce et al. 1998; Devan and White 1999). Moreover, depending on the training protocol, the strategies can be switched immediately after the appearance or disappearance of relevant sensory cues (Devan and White 1999), or learned progressively across trials to prefer one type of cues over another (Pearce et al. 1998). In summary, although these and other behavioral and lesion's studies provide valuable information concerning the influence of sensory cues on behavior and the types of interactions between strategies, the mechanism of the strategy selection is not clear.

In this article, we propose a bio-inspired computational model of selection between response- and place-based strategies applied for navigation in continuous space. This model is based on three key assumptions. The first one is that these strategies are mediated by separate memory systems that can learn independently and in parallel (as in the computational models of, e.g., Guazzelli et al. 1998; Girard et al. 2005; Chavarriaga et al. 2005; Daw et al. 2005). The second assumption is that learning algorithms within the two memory systems are of different types: while response-based strategy relies on a slow and stereotyped "trial-and-error" learning implemented as a temporal-difference (TD) learning procedure, learning in the place-based strategy is fast and flexible and is based on a graph-search algorithm for finding a goal (as in Guazzelli et al. 1998; Girard et al. 2005; Daw et al. 2005). The third assumption is that the selection mechanism is not fixed but continuously updates its estimates of the relative "goodness" of different strategies (as in Chavarriaga et al. 2005; Daw et al. 2005). The novelty of our approach is in the proposed "common currency" allowing the comparison of strategies that use different learning algorithms for reaching the goal. This common currency is defined as the direction of the goal-oriented movement proposed by each strategy. We show below that the selection criterion based on this common currency, is capable of choosing the best among TD-learning and planning strategies and can be used to solve navigational tasks in continuous state and action spaces.

We use our model to reproduce and analyze rat behavior in two experimental protocols in which response- and place-based strategies were shown to interact with each other (Pearce et al. 1998; Devan and White 1999) with the aim of answering the following questions: (i) what is the mechanism of strategy selection that can result in competition and cooperation between strategies across and within experimental trials? (ii) What is the possible selection criterion, i.e., how can the performance of different strategies (with potentially different learning mechanisms) be compared so that the best strategy is chosen to take control over behavior? and (iii) How different types of sensory cues influence strategy selection? The rest of the article is structured as follows: Section 2 describes the model of strategy selection; Sections 3 and 4 describe the results of computer simulations aimed at reproducing animal data; in Sect. 5, we discuss the results of this study in relation to the previous questions and to other available experimental and theoretical studies; Finally, we conclude in Sect. 6 with the outlook on future study.

2 The model

In the model of navigation under this study, response- and place-based strategies are implemented by two "experts", referred to as *Taxon expert* and *Planning expert* in this article. They represent DLS and Hc–PFC memory systems,

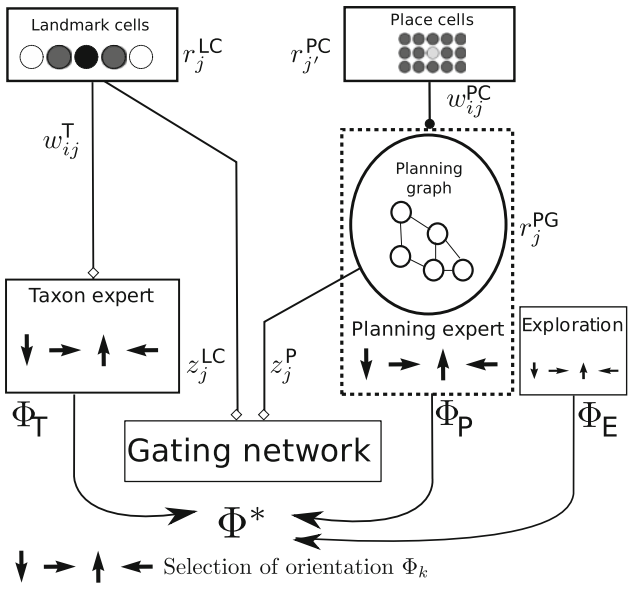

Fig. 1 Model overview (see text for details). *LC* Landmark Cells, *PC* Place Cells, *PG* Planning Graph, *T* Taxon expert, *P* Planning expert, *E* Exploration expert. Φ^* is the direction of the next movement resulting from the selection process

respectively. During navigation, these experts propose a direction for the next movement according to either visual input (Taxon expert) or the estimated location (Planning expert). In addition, the third, *Exploration expert*, proposes a direction of movement randomly chosen between 0 and 2π. The actual movement, performed by the simulated rat (henceforth referred to as "animat"), is determined by the selection module (the gating network) which selects one of the experts to take control over behavior on the basis of previous performance (Fig. 1).

2.1 Taxon expert

The Taxon expert implements response-based strategy in the model. In particular, we consider two kinds of response-based strategies: approaching a visible target (sometimes referred as "beacon learning") and approaching a hidden target marked by a landmark located on a certain distance from it (i.e., "guidance" in terms of O'Keefe and Nadel (1978)). Information about the landmark (or the visible target) is encoded by the activities of N_{LC} Landmark Cells (see Table 1 for parameter values) which code the presence or the absence of the landmark in a particular direction $\phi_i^{LC} = \frac{2\pi i}{N_{LC}}$. The activity of LC i is given by:

$$r_i^{LC} = \exp\left(-\frac{\Delta\Phi_i}{2(\sigma_{LC}/\Delta_{R\to L})^2}\right), \tag{1}$$

where $\Delta\Phi_i = \Phi^L - \phi_i^{LC}$ is the angular distance between the direction of the landmark Φ^L and the cell's preferred direction, and $\Delta_{R\to L}$ is the distance from the animat to the landmark in centimeters (see in, e.g., Brown and Sharp (1995), Touretzky and Redish (1996), for similar modeling of sensory input). The width of the Gaussian centered at the landmark direction increases as the animat approaches the landmark, expressing the fact that the landmark image takes up a larger part of the view field if the animat is close to the landmark.

In the model of our study, the Taxon expert can work in either allocentric or egocentric directional reference frames. The allocentric reference frame is fixed with respect to distal (room) cues and is assumed to be supported by the head direction network involving the anterodorsal nucleus of thalamus (Taube et al. 1990). In this reference frame, the direction to the landmark Φ^L is given with respect to the zero direction that is defined at the first entry to the environment (see Fig. 2) and remains fixed thereafter. In the second, egocentric reference frame, Φ^L is given relative to the zero direction that coincides with the current gaze direction of the animat.

The motor response of the Taxon expert to the landmark stimulus is encoded by $N_{AC} = 36$ *Action Cells* (AC), so that each AC i receives input from all LCs and codes for movement direction $\phi_i^T = \frac{2\pi i}{N_{AC}}$ in the corresponding reference frame. Its activity represents the value of moving in the corresponding direction and is computed as follows (note that superscript T in the following text denotes Taxon expert and not matrix transposition):

$$a_i^T(t) = \sum_{j=1}^{N_{LC}} r_j^{LC}(t) w_{ij}^T(t). \tag{2}$$

The activity in the AC population is interpreted as a population code for the continuous direction Φ^T of the next movement of the animat, proposed by the Taxon expert (Strösslin et al. 2005; Chavarriaga et al. 2005):

$$\Phi^T(t) = \arctan\left(\frac{\sum_i a_i^T(t)\sin(\phi_i^T)}{\sum_i a_i^T(t)\cos(\phi_i^T)}\right). \tag{3}$$

Learning of the weights is performed by the TD-based Q-learning algorithm (Sutton and Barto 1998). We consider the activity $a_i^T(t)$ of an AC i to be the Q-value of the corresponding state–action pair, giving rise to the following formula for the weight update (Strösslin et al. 2005; Chavarriaga et al. 2005):

$$\Delta w_{ij}^T = \eta\delta^T(t) e_{ij}^T. \tag{4}$$

where η is the learning rate, $\delta^T(t)$ is the reward prediction error, and e_{ij}^T is the eligibility trace. The reward-prediction error is defined as the difference between the current and previous estimates of the discounted future reward (Sutton and Barto 1998):

$$\delta^T(t) = R(t+1) + \gamma \max_a a_i^T(t+1) - a^T(t), \tag{5}$$

Table 1 Parameters of the experts

Name	Value	Description
Taxon expert and gating network		
N_{LC} [1]	100	Number of Landmark Cells
σ_{LC} [1]	27.5°	Normalized landmark width
$N_{\mathrm{AC}}^{\mathrm{T}}$ [2]	36	Number of action cells
σ^{T} [2]	22.5°	Standard deviation of the generalization profile
η [3]	0.001	Learning rate
λ [2]	0.76	Eligibility trace decay factor
γ [2]	0.8	Future reward discount factor
ξ [2]	0.01 / 0.05	Learning rate of the gating network (depending on the experiment)
Planning expert		
θ^{PC} [3]	0.3	Activity threshold for place-cells node linking
θ^{P} [3]	0.3	Activity threshold for node creation
α [3]	0.7	Decay factor of the goal value
N_{PC} [4]	1681	Number of simulated Place Cells
σ_{PC} [4]	10 cm	Place field size

[1] Set to give sufficient detailed representation; [2] adapted from Chavarriaga et al. (2005); [3] hand-tuned; [4] set to give a sufficient overlap between place fields

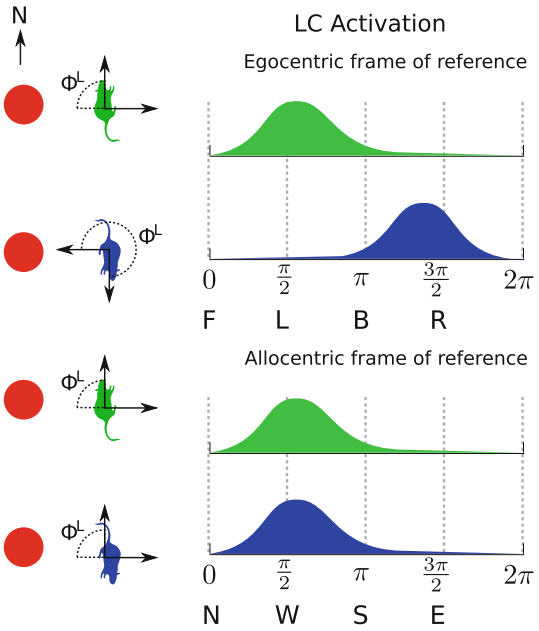

Fig. 2 Internal representation of the landmark (*black dot*) in the egocentric (*top*) and allocentric (*bottom*) spatial reference frames. In the egocentric reference frame, the landmark seen by the animat oriented toward north (marked by *light grey*) or toward south (marked by *dark grey*) will be represented by highly active Landmark Cells at egocentric directions $\Phi^{\mathrm{L}} = \pi/2$ (i.e., on the left side relative to the animat's head direction) and $\Phi^{\mathrm{L}} = 3\pi/2$ (i.e., on the right side), respectively (see Eq. 1). In the allocentric reference frame, the landmark will be represented by highly active cells at the allocentric direction $\Phi^{\mathrm{L}} = \pi/2$ in both cases, since the landmark is located in the western direction from the animat (here, the north direction was chosen as the zero direction of the allocentric reference frame). *F* front, *L* left, *B* back, *R* right; *N* north, *W* west, *S* south, *E* east

where $R(t)$ is the reward delivered at time t, $0 < \gamma < 1$ is the future reward discounting factor, and $a^{\mathrm{T}}(t)$ is the Q-value of the action performed at time t, estimated by the Taxon expert. The eligibility trace e_{ij}^{T} in Eq. 4 speeds up learning by remembering the state–action pairs experienced in the past:

$$e_{ij}^{\mathrm{T}}(t+1) = r_j^{\mathrm{LC}}(t) r_i(t) + \lambda e_{ij}^{\mathrm{T}}(t), \tag{6}$$

where $\lambda < 1$ is the eligibility trace decay rate, $r_j^{\mathrm{LC}}(t)$ is given by Eq. 1 and r_i^{AC} is given by:

$$r_i^{\mathrm{AC}}(t) = -\exp\left(\frac{\phi_i^{\mathrm{T}} - \Phi^{\mathrm{T}}(t)}{2\sigma^{\mathrm{T}2}}\right). \tag{7}$$

This term represents the activity of action cells in the *generalization phase* (Strösslin et al. 2005) and allows the actions which are close to the actually performed action Φ^{T} (Eq. 3) to update their weights in the same direction. The use of a generalization phase for action learning, together with the use of Eq. 3 for action selection results in the ability of the Taxon expert to work in a continuous action space (Strösslin et al. 2005).

We note that the learning algorithm described above does not depend on the spatial reference frame (i.e., allocentric or egocentric, see Fig. 2) that is used. The information about the reference frame is implicitly encoded by the landmark information. However, the learned behavior of the animat in some tasks can be different, depending on what reference frame is used as illustrated in the results (Sect. 3.2.5).

The calculation of the reward-prediction error (Eq. 5) and the corresponding weight update (Eq. 4) are performed on

each time step independently from the identity of the expert (i.e., Taxon, Planning or Exploration) that generated the last action. Moreover, reward signal $R(t)$ is shared between all the experts at each time step. Therefore, goal-oriented actions performed under the control of, e.g., the Planning expert, help the Taxon expert to adjust its weights. This way, the *cooperation* between strategies is implemented in the model, in addition to the *competition* between strategies, governed by the selection network (see Sect. 2.3 below for the competitive selection algorithm).

2.2 Planning expert

The Planning expert uses a simple graph-search algorithm to find the shortest path to the goal (Martinet et al. 2008). During an unrewarded *map building* phase, the Planning expert builds a graph-like representation of space based on the activities of simulated *Place Cells*. During a reward-based *goal planning* phase, this representation is used to plan and execute goal-directed path. Since extra-maze cues are stable in the experiments that we will simulate, we use a simple model of Place Cells as described later (see Arleo and Gerstner 2000; Sheynikhovich et al. 2009 for more detailed models of Place Cells that integrate information from distal cues and path integration). The population of Place Cells in our model is created before the learning is started, and the activity of place cell j is given by

$$r_j^{\mathrm{PC}} = \exp\left(-\frac{\Delta^2_{\mathrm{A}\to j}}{2\sigma^2_{\mathrm{PC}}}\right), \tag{8}$$

where $\Delta_{\mathrm{A}\to j}$ is the distance between the animat and the center of firing field of place cell j (i.e., place field), and σ_{PC} is the width of the place field. Place field centers are distributed uniformly in the environment.

Given the Place Cells activity, the *Planning Graph* is built during unrewarded movements by the following algorithm. When a new node N_i is created, it is connected to place cell j with connection weights w_{ij}^{P}:

$$w_{ij}^{\mathrm{P}} = r_j^{\mathrm{PC}} \mathcal{H}\left(r_j^{\mathrm{PC}} - \theta^{\mathrm{PC}}\right), \tag{9}$$

where $\mathcal{H}(x) = 1$ if $x > 0$, $\mathcal{H}(x) = 0$ otherwise. The activity r_i^{P} of node i is then computed by

$$r_i^{\mathrm{P}} = \sum_j r_j^{\mathrm{PC}} w_{ij}^{\mathrm{P}}. \tag{10}$$

A new node is added on each time step unless at least one existing node is active above threshold θ^{P}. The overlap between PCs, threshold values θ^{PC}, and θ^{P} have been chosen to guarantee that, when the condition for the creation of a new node is met (i.e., no node activity above θ^{P}), there is always at least one PC, whose activity is above θ^{PC}. Thus, any newly created graph node has at least one connection weight to the PCs that is non-zero.

A link between nodes N_i and N_j stores the allocentric direction of movement required to pass from one node to the other:

$$\Phi^{\mathrm{P}}(t) = \widehat{\vec{x}\,\overrightarrow{N_i N_j}}, \tag{11}$$

where x is the zero angle of the allocentric reference frame. This link is created only when the animat travels between two nodes, with no intermediary node having already been present. This means that when a new node is created, it is already connected to the node previously visited by the animat. Therefore, a node i will be connected to the node j if and only if there is no node k such that

$$\widehat{\overrightarrow{N_i N_j}\,\overrightarrow{N_i N_k}} < \epsilon \quad \text{and} \quad \|\overrightarrow{N_i N_k}\| < \|\overrightarrow{N_i N_j}\|, \tag{12}$$

where ϵ is dependent on the moving and rotation speeds of the animat. This insures that graph nodes are only connected to their closest neighbors.

Given the Planning Graph, the optimal path to the goal is determined by the activation–diffusion mechanism (Burnod 1991; Hasselmo 2005), based on the Dijkstra's algorithm for finding the shortest path between two nodes in a graph (Dijkstra 1959). More specifically, during goal planning, the Planning expert first determines its location using a *position value* and then calculates the direction toward the goal using *goal value*. The position value corresponds to the activity r_i^{P} of the node (Eq. 10). The goal value $G_i = 0$ when no goal position is known. In this case, the strategy proposes a random movement direction among the different possible actions from the current node. In contrast, when the goal position is found (using the actions generated by any expert), the goal value of the closest (goal) node is set to $G_{i^*} = 1$, and is propagated to all the adjacent nodes, decreased by a decay factor $\alpha < 1$. The goal value G_i of a node i of distance n from the goal node (measured as the number of nodes between the goal node and the node i) is given by $G_i = \alpha^n$. The next movement direction is given by the link to the adjacent node with the highest goal value.

2.3 Strategy selection

During goal learning, the model has to select out of the three experts, Taxon, Planning, and Exploration experts (T, P, and E, respectively), which one takes control over behavior, i.e., chooses the next action. The gating network learns to select experts on the basis of the "common currency" defined as the direction of movement proposed by each expert. After learning, the expert that proposes directions of movements that are closest to the true direction to the goal is considered the best at each time step. We use only three experts at present, but the selection network can work with any number of experts

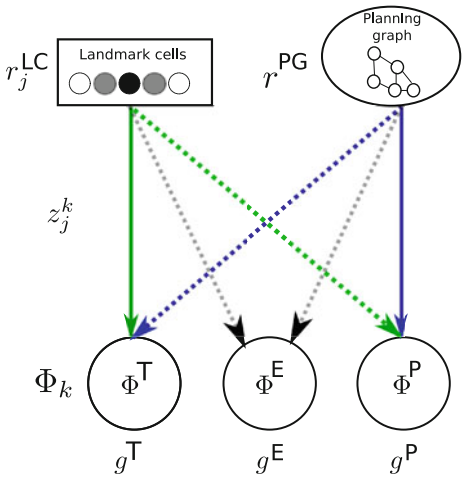

Fig. 3 Gating network. The inputs of the Taxon and Planning experts (LC and PG) are linked to the units in the gating network. The gating values g^k are weighted sums of the input values r_j with weights z_j^k. One of the three experts is selected according to a winner-take-all scheme

as long as they provide a direction of movement toward the goal as their output.

In the present model, the gating network consists of three units $k \in \{T, P, E\}$, each corresponding to a separate expert. The activity g^k of the unit k is called "gating value" of the corresponding expert. The input to the units in the gating network is provided by the activities of the LC population and the nodes of the Planning Graph (Fig. 3). The gating values g^k are calculated as

$$g^k(t) = \sum_{j=1}^{N_{LC}} z_j^k(t) r_j^{LC}(t) + \sum_{j=N_{LC}+1}^{N_{LC}+N_P} z_j^k(t) r_j^P(t), \tag{13}$$

where z_j^k is the connection weight between the unit k of the gating network and input unit j of the experts. As described in the previous sections, at each time step experts propose candidate directions Φ^k of the next movement. The gating values are used to choose the next movement direction Φ^* to be taken by the animat using a winner-take-all scheme:

$$\phi^k(t); k = \text{argmax}_i(g^i(t)) \tag{14}$$

Similar to the learning in the Taxon expert, the connection weights for the Taxon and Planning gating values are randomly initialized between 0 and 0.01 and adjusted using a Q-learning algorithm. The weight update in this case is given by

$$\Delta z_j^k = \xi^G \delta^G(t) e_j^k(t), \tag{15}$$

where ξ^G is the learning rate of the gating network, and $\delta^G(t)$ is the reward-prediction error:

$$\delta^G(t) = R(t+1) + \gamma \max_k \left(g^k(t+1)\right) - g^{k^*}(t), \tag{16}$$

Here, $R(t)$ is the reward delivered at time t, γ is the future reward discount factor of the gating network, and g^{k^*} is the gating value of the expert, chosen at time step t (i.e., the time step that corresponds to the direction of movement in Eq. 14).

As for the Taxon strategy, the eligibility trace e_j^k of expert k allows the gating network to reinforce the experts selected in the past:

$$e_j^k(t+1) = \Psi(\Phi^*(t) - \Phi^k(t)) r_j^k(t) + \lambda e_j^k(t), \tag{17}$$

where λ is the eligibility trace decay factor. The term $\Psi(\Phi^*(t) - \Phi^k(t))$ can be considered as a discrete version of the action generalization in the Taxon expert, where

$$\Psi(x) = \exp(-x^2) - \exp(-\pi/2). \tag{18}$$

This term insures that the closer the orientation is from the selected one, the higher the corresponding strategy will be reinforced ($\Psi(x)$ is maximum when $x = 0$). In contrast, two strategies that proposed two opposite orientations will have opposite reinforcements. The selection between experts is performed at each time step, unless the Exploration expert is chosen, in which case the chosen orientation is taken during three subsequent time steps. This was done to avoid the animat being stuck in a particular location due to random weight initialization. Since exploration actions are pseudo-random, their weight will *decrease* with learning relative to the weight associated with strategies that direct the animat toward the goal (since the gating network assigns higher weights to strategies that maximize reward). This situation does not change when the weights start to converge, since exploration strategy will not predict rewards better at the end of training; its actions remain always pseudo-random.

3 Simulation I: Experiment of Pearce et al. (1998)

In this experiment, two groups of rats (Control and Hippocampal-lesioned) learned to find the location of a hidden platform in a circular water maze. A visible landmark was located in the pool at a certain distance and allocentric direction from the platform. At the start of an experimental session, the platform and the landmark were moved to one of eight predefined locations in the pool (Fig. 4a), and remained fixed for four trials, after which a new session started. The principal observed results of this experiment (see Fig. 3 in their article) consisted in the observations that (i) both the lesioned and intact rats learned to swim to the hidden platforms at the end of training, and (ii) escape latencies of Hc-lesioned rats were significantly shorter than Control rats in the first trials of intermediate sessions, while they were significantly longer than Control rats in the last trials of each session. From these results, the authors concluded that the intact rats used two competing navigation strategies to locate the goal: a

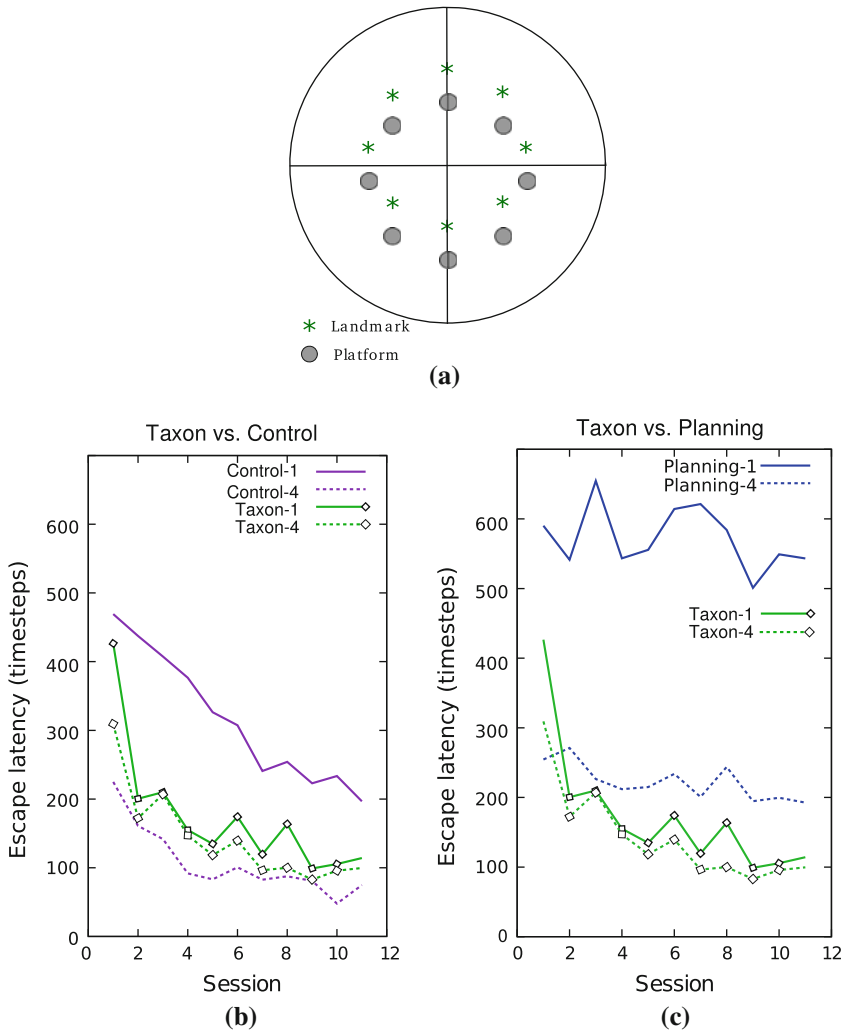

Fig. 4 **a** Experimental setup of Pearce et al. (1998). Mean escape latencies of simulated rats across sessions. **b** Control versus Taxon group. **c** Planning versus Taxon group. *Solid*, and *dotted lines* correspond, respectively to first-trial and last-trial latencies

Hc-dependent strategy that remembered the goal location with respect to distal extra-maze cues; and a Hc-independent strategy (termed "heading vector strategy" by the authors) that remembered the allocentric direction from the landmark to the goal.

3.1 Simulation procedure and data analysis

The simulated water maze, rat, and landmark were represented by circles of 200, 15, and 20 cm in diameter, respectively. The reward location of 10 cm in diameter was always located 20 cm south from the landmark. At the start of a session, the platform and the associated landmark were randomly moved to one of the eight positions, as shown in Fig. 4a. At the beginning of each trial, the animat was placed in one of the four cardinal positions near the wall, with a random initial orientation. The starting locations were pseudo-randomly avoiding two consecutive trials with the same start location. The moving speed of the animat was set to 18 cm/s, with a simulation time step corresponding to 1/3 s. If the animat was not able to reach the platform in 200 s, it was automatically guided to it along a direct path to the target, similarly to the real rats in this experiment. Reaching the goal was rewarded by $R = 1$, and wall hits were punished by $R = -0.5$ (see Eq. 5 and 16).

The intact rats were simulated by a full model (Control group), including Taxon, Planning, and Exploration experts. Two lesion groups were simulated: animats in the Taxon group used only Taxon and Exploration experts, while animats in the Planning group used only Planning and Exploration experts. The Taxon group corresponded to the Hc-lesioned animals of the original experiment. The allocentric version of the Taxon version was used in this simulation (see Sect. 5.3.2).

In all simulations now being discussed, the results were averaged over 100 animats (noise in the system was due to the random initialization of weights and random choice of starting position). Both across and within sessions, performance of Control, Taxon and Planning groups were statistically assessed by comparison of their mean escape latencies—the number of time steps per trial—in the first and the fourth trials

of a session, using signed-rank Wilcoxon test for matched-paired samples. Between-group comparison was performed using a Mann–Whitney test for non-matched-paired samples. Animat behavior was characterized by three measures: *Goal occupancy rate* of a goal location, defined as the number of times the animat visited a rewarded zone, divided by the total trajectory length; *Goal selection rate* of an expert, calculated as the number of times this particular expert was chosen within a square zone of 0.4 m^2 around the goal, divided by the total number of times the animat visited this zone; *Trial selection rate* of an expert, defined as the number of times the expert was selected over the total length of the trajectory.

The competitive interaction between strategies was estimated by the negative correlation (Pearson's product-moment coefficient) of their selection rates x and y calculated as $\rho_{x,y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$, where σ_{xy} is the covariance, and σ_x, σ_y are the standard deviations of the selection rates x and y, respectively.

3.2 Simulation results

3.2.1 Learning across and within sessions

Both the simulated Control and Taxon groups were able to learn the location of a hidden platform, as shown by the decrease of their escape latencies (Fig. 4b; $P < 0.001$ for all groups). Moreover, in contrast to the Taxon group, animats in the Control group decreased significantly their escape latencies within all the sessions (Control-1 vs. Control-4 in Fig. 4b).

A comparison of two simulated lesion groups (Taxon and Planning groups) shows that the Taxon expert was responsible for decreasing escape latencies across sessions, while the place-based expert was responsible for learning within sessions (Fig. 4c). Moreover, the Control group found the platform more quickly in the fourth trials (dotted line in Fig. 4b) than both the Taxon and Planning groups (dotted lines in Fig. 4c), suggesting that the two strategies cooperated during learning. This was also assessed by their current goal occupancy rate that increases in fourth trials (Fig. 5a).

Similar to real rats, simulated Control group had greater escape latencies than Taxon group in the first trials (Fig. 4b). Pearce et al. (1998) suggest that this might be explained by the preferential use of the Hc-based strategy at the end of a session, so that, at the beginning of a new session (when the platform has moved to a new location), this strategy led the animal to the previous (thus wrong) platform location. In order to check whether this is the case in our model, we calculated goal occupancy rates near previous and current goal locations for simulated Control and Taxon groups. The results show that indeed, the Control group had a significant bias toward the previous goal location on first trials (Fig. 5b, first trials), while this bias disappeared after the Planning expert had learned the new goal location (Fig. 5b, fourth trials). The reason for this is that the Planning expert of our model was not able to notice that platform and landmark have been moved to a new location at the start of a session, in contrast to the Taxon expert.

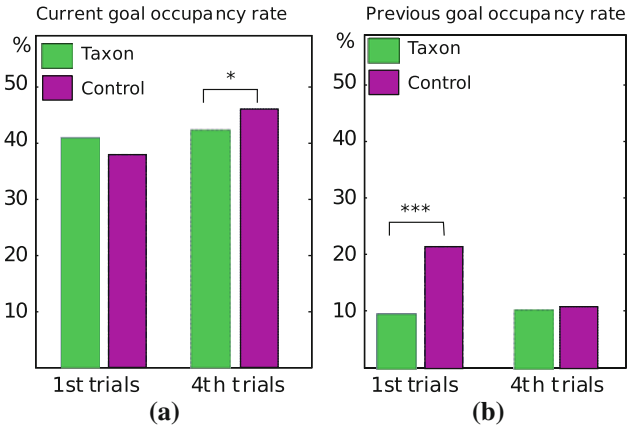

Fig. 5 Occupancy rates for **a** the current and **b** previous goal locations for simulated Taxon and Control groups. *** and * correspond respectively to significance levels $P < 0.001$ and $P < 0.05$

Thus, the overall performance of the model in this task is consistent with that reported by Pearce et al. (1998). The advantage of the modeling approach applied here is that we can go further in our analysis of behavior and explore the interactions between behavioral strategies *within* experimental trials. Such an analysis is usually hard to perform in animal experiments like that of Pearce et al. (but is possible for simpler tasks, like e.g., Hamilton et al. 2004). Such a complementary analysis allows us to get insights into (i) the importance of different types of sensory cues for different strategies and (ii) competitive and cooperative interactions between trials across and within experimental sessions.

3.2.2 Influence of sensory cues

In order to analyze the importance of landmark versus spatial cues on learning, we compared the synaptic weights between the connections from Landmark Cells (that encode the landmark) and nodes of Planning Graph (that encode location) to the units of the gating network, which encode the two strategies in the model. The observed increase in the average weights for all connections suggests that all types of cues played a role in the selection process (Fig. 6). However, weights from Landmark Cells to both the Taxon and Planning gating units grew significantly faster with learning, than those from Planning Graph nodes ($P < 0.01$, see caption of Fig. 6). These results suggest that, in our model, the landmark exerted progressively stronger influence on strategy selection than spatial cues, which is consistent with the fact that this task could be solved only by paying attention to the landmark. Nevertheless, the spatial cues were also learned, although

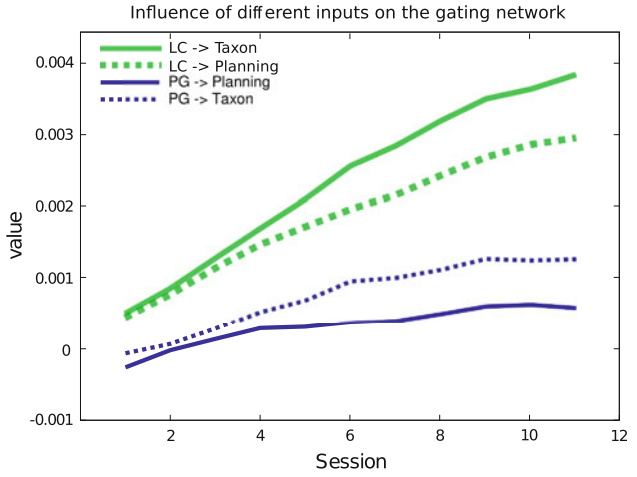

Fig. 6 Evolution of the average synaptic weights between inputs of the gating network and gating units of different strategies. *Thick lines* represent *straight links* (LC → Taxon, PG → Planning). *Dotted lines* represent cross links (LC → Planning, PG → Taxon). A linear regression test on these slopes indicates that LC → Taxon weights grow 5.4 times faster than PG → Taxon weights. Accordingly, LC → Planning weights grow 2.3 times faster than PG → Planning weights

with a smaller rate, and so could influence selection when Planning expert becomes more efficient.

3.2.3 Competition between strategies across experimental sessions

Next, we analyzed the competitive interaction between experts in the Control group across training sessions by comparison of their goal and trial selection rates. Pearce et al. (1998) suggest that, at the beginning of each session, the place-based strategy was in competition with the heading-vector strategy, the latter being the winner of the competition by the end of training. We checked whether our model is consistent with this hypothesis.

At the start of a new session, the Planning expert was not able to detect the change in the platform location and hence its goal selection rate did not change significantly from earlier to later sessions (Fig. 7a, first trials). Accordingly, the first trial selection rate of the Planning strategy did not change significantly across sessions (Fig. 7b). In contrast, the Taxon expert learned to track the changes in landmark position, as suggested by the progressive increase of its trial selection rate across experimental sessions (Fig. 7b, first trials), and by the significant increase in its goal selection rate in the later sessions relative to earlier sessions (Fig. 7c). The competitive interaction between the Taxon and Planning experts is illustrated by the typical trajectory of the simulated animal at the beginning of a session (Fig. 8a). The Planning expert led the animat toward the previous platform location, while the Taxon expert led it toward the current one.

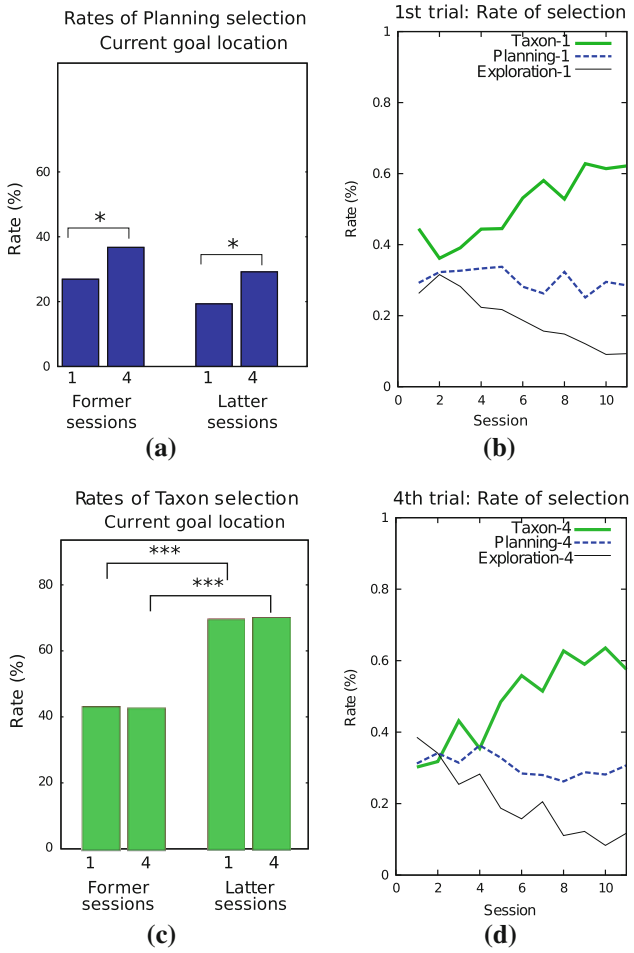

Fig. 7 **a** Selection rates of the Planning expert near the current goal location, across and within sessions. **b** Strategy selection rates across sessions in first trials. **c** Taxon strategy selection rate near the current goal location. **d** Strategy selection rate across the sessions in fourth trials. *** and * correspond respectively to significance levels $P < 0.001$ and $P < 0.05$

Interestingly, the decrease in the trial selection rate of the Exploration expert was almost opposite in magnitude to the increase in the Taxon selection rate (correlation coefficient $r = -0.96$). This result suggests that the preferential use of the Taxon strategy at the end of training corresponds to a decrease in exploratory behavior, rather than a decrease in place-based strategy (Fig. 7b).

3.2.4 Cooperation between strategies within a session

As shown above, the competitive interactions between Taxon and Planning strategies were due to the fact that these two strategies encoded different goal locations at the start of a session. However, this situation changed by the end of session when both strategies had learned the true goal location. In both early and late sessions, the Planning expert was selected significantly more often near the current goal location in

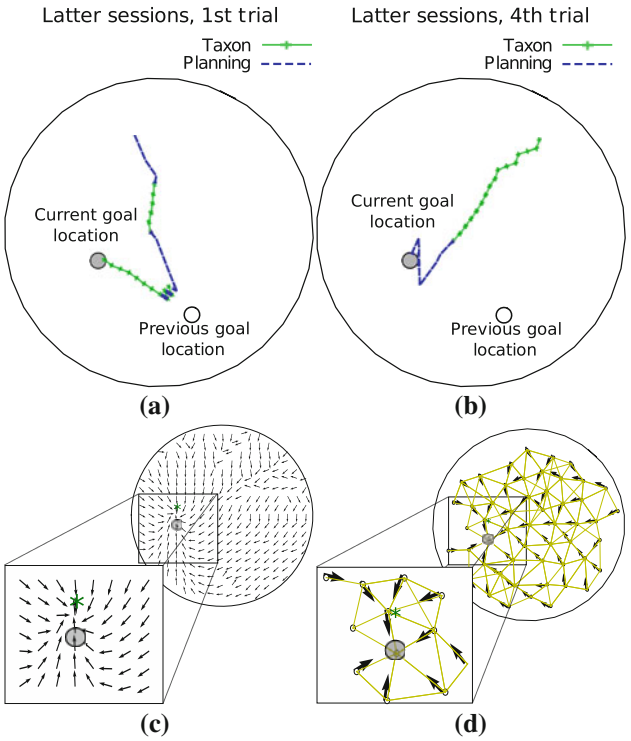

Fig. 8 Control group **a** example of typical trajectories in last sessions of the first trial, **b** example of typical trajectory in the last sessions of the fourth trials and the associated navigational maps around the goal location of **c** Taxon and **d** Planning strategies

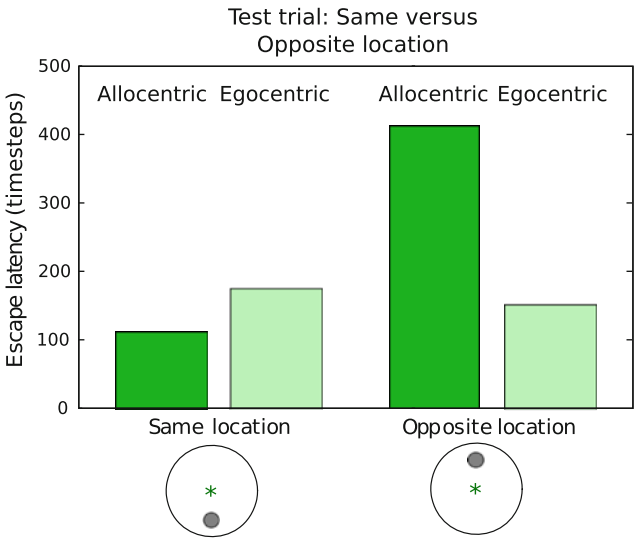

Fig. 9 A correspondence between the allocentric Taxon strategy in the model and the heading-vector strategy (Pearce et al. 1998). The plot shows the mean escape latency to find the platform hidden in the same location relative to the landmark as during training (same location), or in the location opposite to it (opposite location). Contrary to the egocentric Taxon expert, the allocentric Taxon expert had difficulty in finding the platform in the opposite location, since it "remembers" the allocentric direction from the landmark to the hidden goal

fourth trials than in first trials (Fig. 7a), whereas Taxon expert was selected near the current goal location as much often in fourth trials as in first trials in both early and late sessions (Fig. 7c). The increase in Planning selection rate near goal, without provoking a decrease of the Taxon selection rate, and superior performance of Control group over other groups in the fourth trials (Fig. 5a) suggests a cooperative interaction between both experts. Such a cooperative interaction is illustrated by a typical trajectory in the fourth trial (Fig. 8b). Here, both strategies led to the correct goal location and the choice of a particular strategy depended on the accuracy of the corresponding expert at different locations along the trajectory. Examples of navigational maps of the two experts near the goal location are shown in Fig. 8c, d. In these maps, arrows corresponding to the learned directions of movement for each sample location (Taxon expert) or for each spatial node (Planning expert), show that the Taxon expert points southward the landmark, and the Planning expert toward the platform location.

3.2.5 Allocentric Taxon strategy as a heading-vector navigation

In the simulation shown above, we used an allocentric version of the Taxon expert to reproduce the rat behavior attributed by Pearce et al. (1998) to heading-vector navigation. They defined the heading-vector strategy as follows: rats "might use a heading vector that specifies the direction and distance of the goal from a single landmark." Here, we show that the allocentric Taxon expert suits well this definition.

In order to demonstrate that the allocentric taxon strategy in the model is similar to the "heading-vector" strategy observed in rats, we performed behavioral test similar to that used in the original experiment. After training the Taxon group in 11 sessions of the main experiment, the landmark was placed at the center of the pool. In the case of half the number of the animats in the simulated Taxon group, the platform was located south of the landmark and at the same distance as before, while for the other half, the platform was located north of the landmark. We compared the performance of the allocentric and egocentric versions of the Taxon expert in the model. Similar to the Hc-lesioned animals, animats with allocentric Taxon expert for which the platform was located north of the landmark took significantly longer to locate the platform than the other group (Fig. 9). This is explained by the fact that the allocentric taxon strategy relies on the remembered allocentric direction from the landmark to the goal, while the egocentric taxon strategy cannot use this information, and hence searches randomly around the landmark (see Sect. 2.1). From these results we conclude that the allocentric Taxon expert is a suitable model of the heading-vector strategy observed by Pearce et al.

In summary, our results support the hypothesis of Pearce et al. (1998) that, at the beginning of the training sessions, place- and response-based strategies were in competition with each other. However, on the basis of results of this study, we propose that, at the end of a session, a cooperation between strategies takes place. In addition, we propose that the improvement of the rat performance by the end of training is not due to the decrease in the use of place-based strategy, but rather due to the decrease in the number of exploratory actions. We stress here that in the model described, the trade-off between exploration and exploitation is not fixed, but learned during training (see Sect. 5).

4 Simulation II: Experiment of Devan and White (1999)

In this experiment, sham-operated, fornix-lesioned and DLS-lesioned groups rats were trained for nine days to remember the location of a platform in a water maze. On days 3, 6, and 9 the platform was hidden, whereas it was visible on the other days. During a competition test on day 10, the visible platform was placed in a novel location (Fig. 10a).

Four principal findings from the original experiment were related to the issue of interaction between place- and response-based strategies (see Fig. 2 in their article). First, sham-operated rats, rats with fornix/fimbria lesions and rats with DLS lesions were equally fast in learning the visible platform location, suggesting that either strategy can be used to approach a visible goal. Second, rats with fornix/fimbria lesions were slower than both sham-operated and DLS-lesioned rats during the hidden platform sessions, suggesting that Hc-dependent strategy, and not the DLS-dependent strategy, is required to locate the hidden platform. Third, on the competition test, rats with fornix/fimbria lesions escaped faster from the pool than either sham-operated or DLS-lesioned groups, suggesting a competition between the two strategies. Fourth, the authors identified two groups of sham-operated animals during the final test day: "place-responders" were approaching the place where the hidden platform was in the previous trial, discarding information from the visible platform in a new place; "cue-responders" headed toward the visible platform and were not biased by the hidden platform location in the previous trials.

4.1 Simulation procedure and data analysis

The experimental setup was similar to that used in Simulation I, except that the diameter of the water maze was set to 172 cm to be consistent with the original protocol. On days 1, 2, 4, 5, 7, 8, the visual landmark 10 cm in diameter (representing the visible platform) was placed into the center of the southwest quadrant of the environment (its position coincides with the reward zone). On days 3, 6, and 9, the landmark was absent, but the reward zone remained in the same location. On day 10, the landmark together with the reward zone were moved to the center of the northeast quadrant of the environment. Starting positions were chosen as in Simulation I. On the competition test the starting position equidistant from both landmark locations was chosen.

Sham-operated, fornix-lesioned and DLS-lesioned groups were respectively simulated by the Control, Taxon and Planning groups as in Simulation I. In this simulation we used the egocentric version of the Taxon expert (see Model and Sect. 5.3.2). The same statistical tests as in Simulation I were used to assess learning.

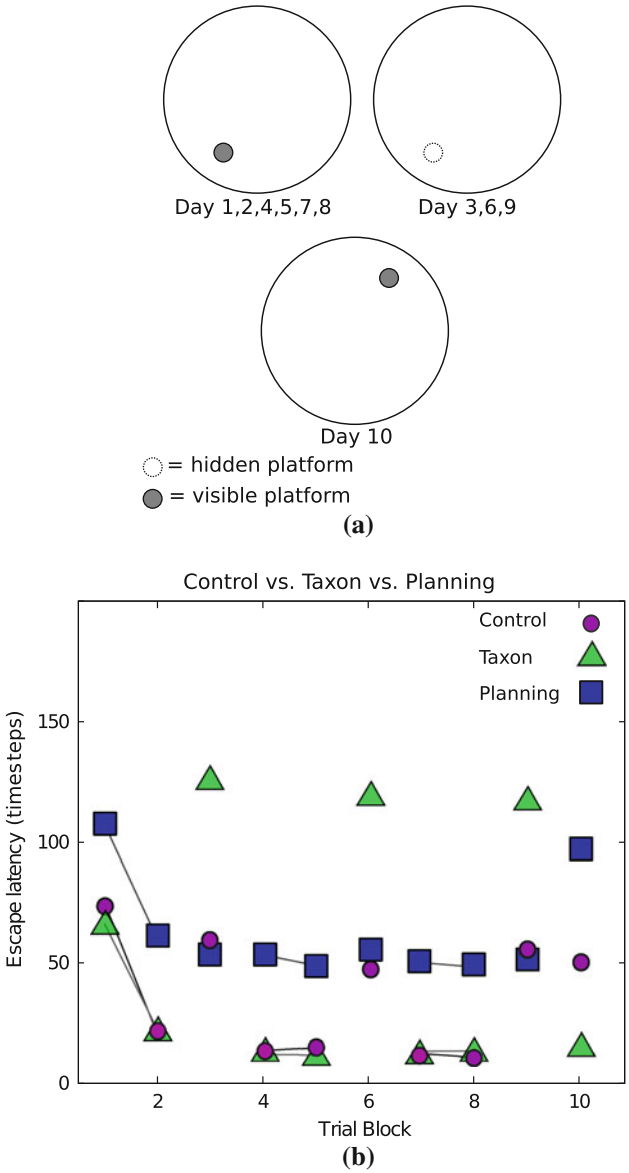

Fig. 10 **a** Protocol of the experiment. **b** Mean escape latencies of simulated rats in Control, Taxon, and Planning groups across sessions with visible (connected plot, days 1, 2, 4, 5, 7, and 8) and hidden (days 3, 6, and 9) platform. Competition test was conducted on day 10 (see text)

4.2 Simulation results

4.2.1 Parallel learning of navigational strategies

When the visible landmark was signaling the platform location, all groups of animats were successful in learning the goal location (Fig. 10b, trial blocks 1, 2, 4, 5, 7, and 8). The Planning group was longer than Taxon and Control groups. In this model, this is a consequence of the fact that the platform location did not usually coincide with a graph node, resulting in the lower precision of the Planning Graph compared to the visual input and elevated use of the Exploration expert. The performance of Control and Taxon groups was not different, suggesting that in this case, the behavior of the Control animats was controlled primarily by the Taxon expert.

When the reward location was not signaled by the landmark, the Taxon group had significantly longer escape latencies, that did not decrease with training, similarly to the rats with fornix/fimbria lesions (Fig. 10, trial blocks 3, 6, 9). The performance of the Control group was not different from that of Planning group, suggesting that, in these trial blocks, the behavior was controlled by the Planning expert.

On the competition test, animats from the Taxon group were significantly faster than those from either Control and Planning groups in reaching the new platform location ($P < 0.001$, Fig. 10, trial block 10). In addition, Control group was significantly faster than Planning group, whose performance did not differ from that in the first trial. This last difference was not observed in the original experiment, possibly due to the fact that DLS-lesions in rats may have spared some ability to approach a visible target moved to a new position, whereas our animats in the Planning group were not able to do so. Nevertheless, these results are consistent with the finding of Devan and White (1999) that rats with fornix/fimbria lesions performed significantly better on the competition test than both the Control and DLS-lesioned groups.

Taken together, these results show that our selection model is consistent with the rat behavior observed in this experiment. Similar to the analysis performed in Simulation I, in the next section, we focus on the influence of visual cues and on analysis of strategy interaction.

4.2.2 Influence of sensory cues

The evolution of the synaptic weights in the gating network reflected the irrelevance of the Taxon expert for the trials in which the goal is hidden (Fig. 11). This was expressed by the progressive decrease of the connection weights between spatial cues and the gating unit corresponding to the Taxon strategy. This is in marked contrast with the weight evolution in Simulation I (Fig. 6), where both types of cues were present throughout training and could be both used to find the goal.

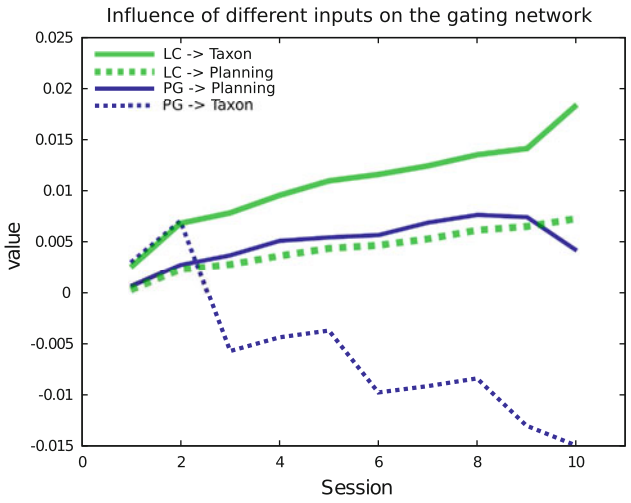

Fig. 11 Synaptic weights of the gating values in the gating network in Control group. *Thick lines* represent straight links (LC → Taxon, PG → Planning). *Dotted lines* represent cross links (LC → Planning, PG → Taxon)

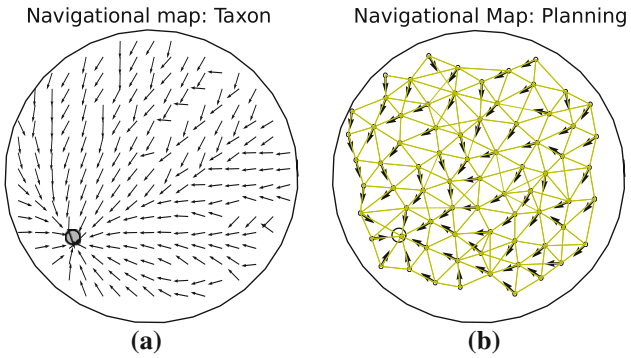

Fig. 12 Navigational maps of **a** Taxon and **b** Planning experts at the end of the trial blocks 8 and 9

4.2.3 The absence of cooperation between strategies during training

During training, both the Taxon and Planning experts learned to approach the fixed goal location. This is illustrated by the navigational maps learned by the two experts (Fig. 12). It can be observed that the map learned by the Taxon expert was more accurate than that of the Planning expert, due to the fact that in this experiment goal location coincided with the landmark. Hence, no cooperation with the Planning expert was necessary in this case. Indeed, trial selection rates of different experts show that the Taxon expert clearly controlled the behavior when the goal was visible (Fig. 13a, b).

In contrast, during trial blocks in which the goal was hidden, the Planning expert was progressively more selected than the Taxon expert (Fig. 13a, b). In addition, the role

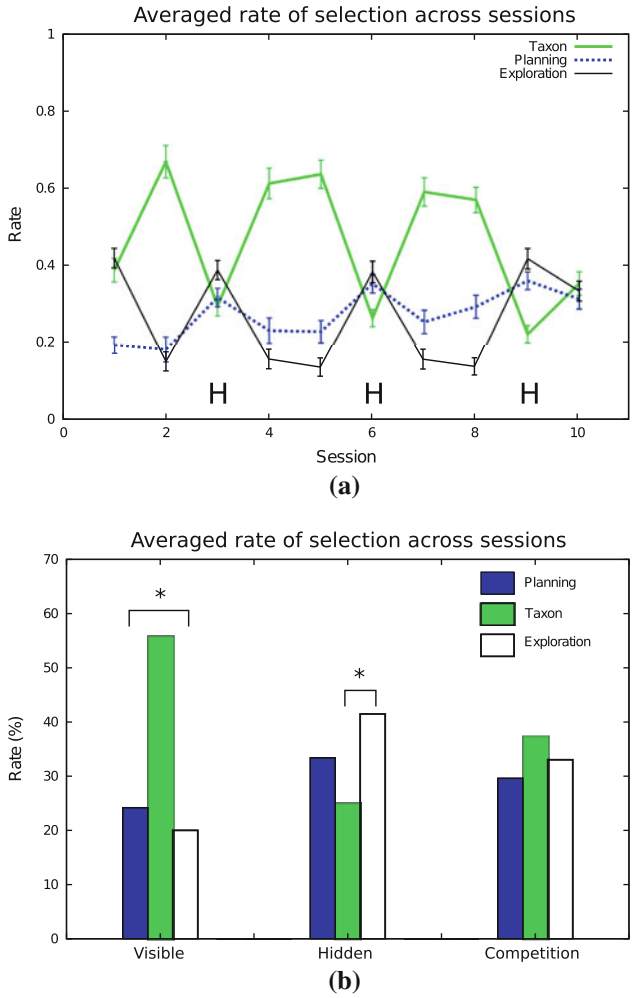

Fig. 13 **a** Selection rates of the three experts during training and competition test in Simulation II (H: Hidden Platform). **b** Summary plot, showing the average rate of selection rate of different experts during trial blocks with visible goal, hidden goal and during competition test

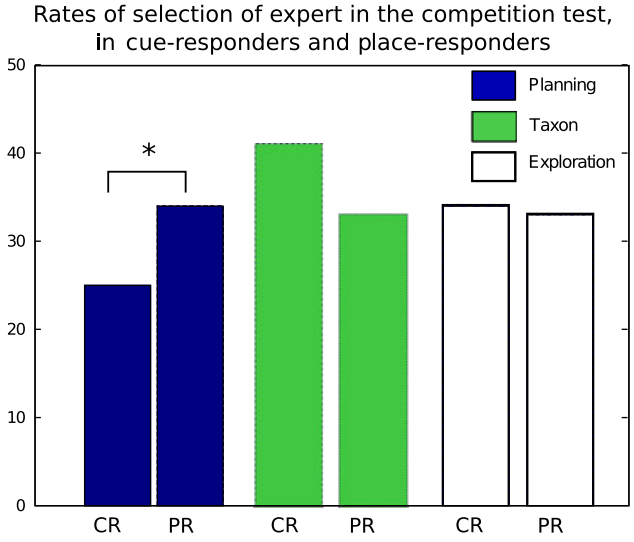

Fig. 14 Rates of selection of experts of cue-responders (CR) and place-responders (PR) in the competition test

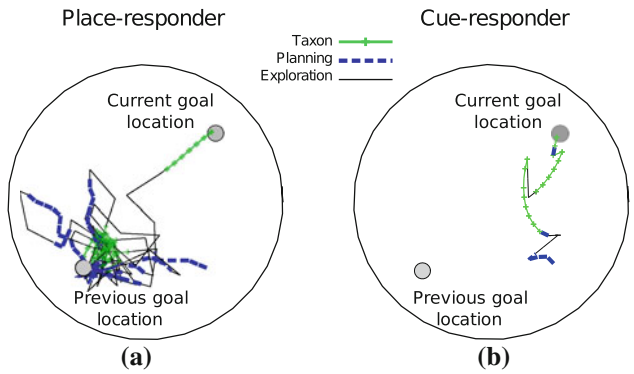

Fig. 15 **a**, **b** Typical trajectories of animats labeled as **a** "place-responders" and **b** "cue-responders"

of exploratory behavior was more prominent in these trials, compensating the relative inaccuracy of the Planning Graph.

4.2.4 Competition between strategies during test

In the competition test, the simulated Control group was able to select a cue-based strategy to reach the goal location, as suggested by escape latencies (Fig. 10) and the selection rate of the Taxon expert (Fig. 13b). However, a significantly better performance of the Taxon group in the competition test (Fig. 10) and a higher selection rate of the Taxon expert during training with visible goal (Fig. 13b) suggests that competition with other strategies slowed down the Control group relative to the Taxon group during the test.

Using the same labeling scheme as Devan and White (1999), Control animats could also be classified into "cue-responders" (59%) and "place-responders" (41%). This division qualitatively reproduced the division of Control rats into both groups of the original experiment (4 "cue-responders" and 6 "place-responders" over 10 animals). Analysis of the trial selection rates of Taxon and Planning experts showed that indeed, in the group of "place-responders" the Planning expert was selected significantly more often than for the group of "cue-responders" ($P < 0.05$), In contrast, in the group of "cue-responders," the Taxon expert was selected more often (although the difference does not reach the significance level, $P = 0.05$, Fig. 14a, b). The observation of typical trajectories of place- and cue-responders shows that place-responders were stuck near the previous goal location during competition test, while cue-responders went almost straight to the visible goal (Fig. 15).

In summary, these results suggest that the proposed selection criterion is flexible enough to deal with rapid strategy switches required when environmental cues drastically change. The Taxon expert in our model learned navigational

maps that were more accurate than those of the Planning expert. The limited number of nodes of the Planning Graph was compensated by the high selection rate of the Exploration expert in the sessions with hidden goal (Fig. 13a). The role of the Exploration expert in our model was to find the exact goal location in an approximate goal area signaled by the "cognitive map" (represented by the Planning Graph), rather than to update the map, as is usually proposed (O'Keefe and Nadel 1978).

5 Discussion

We presented a computational model of switching between cue-guided and place-based strategies in the water maze. The main novel property of this model is that it is capable of learning to select between cue-guided and place-based strategies that use different learning algorithms and spatial reference frames to locate a goal. The place-based strategy uses a graph-search algorithm to find the shortest path to the goal. The graph is learned online using the activities of simulated Place Cells that encode spatial location of the animat in an allocentric reference frame. The cue-guided strategy uses a TD learning rule to approach either a visible goal, encoded in an egocentric reference frame; or a hidden goal marked by a landmark, encoded in an allocentric directional reference frame. The strategy selection is performed by a gating network that learns to predict, using a simple TD-learning rule, the most successful strategy, on the basis of the direction of movement that each expert offers at each time step, given all current sensory inputs.

The model was tested in two simulated water-maze tasks designed to investigate interactions between place- and response-based strategies in rats. Owing to the separation between cooperative (during action learning) and competitive (during action selection) interaction between strategies in the model, we were able to assess the relative contribution of different strategies within, as well as across experimental trials. The sections hereafter shall aim at answering the questions raised in the introduction.

5.1 Strategy selection mechanism

5.1.1 Relation to other models

Several models of strategy switching based on the theory of parallel memory systems were proposed earlier (Guazzelli et al. 1998; Daw et al. 2005; Girard et al. 2005; Chavarriaga et al. 2005). In the model of Guazzelli et al. (1998), the orientations proposed by egocentric taxon and allocentric planning strategies are, respectively, determined by current affordances and cognitive knowledge. The final movement is computed as a sum of these orientations that hand-tuned parameters adapt to the situation. A similar selection is also made in the basal-ganglia loops model of Girard et al. (2005). In these models, strategy switches occur in a set of situations a priori chosen by the modeler. In our earlier study (Chavarriaga et al. 2005; Dolle et al. 2008), the strategy-selection network is adaptive, but it is able to select only between strategies that use TD learning to learn the task. Indeed, the selection network uses TD reward-prediction error as a measure of success of different strategies and hence is not able to deal with other goal-navigation algorithms such as planning. Reinforcement learning framework is also used in the model of Uchibe and Doya (2005) to select between two navigational strategies, but does not handle strategies that are not learned by RL. Finally, the model of action selection in an operant conditioning (Daw et al. 2005) proposes another interesting mechanism of selection depending on the relative uncertainty of different experts. However, in this model, the tree-based computations performed by the experts only allow the model to work with rather small state spaces, and hence cannot be applied to navigation in continuous space. The advantage of the selection criterion proposed in this study is that it permits comparison between experts that use different learning rules and scales well with increasing number of exerts.

5.1.2 The role of random exploration

In the above model, exploration is implemented as a separate "strategy," i.e., during goal learning, it is chosen when its gating value is the highest among the gating values of all the strategies. It means that the need for exploring novel actions is learned during training and can depend on sensory input. This is in contrast to standard reinforcement learning algorithms in which exploration is chosen according to a predefined stochastic scheme. For example, Arleo and Gerstner (2000) and Chavarriaga et al. (2005) use an ϵ-greedy scheme, in which novel actions are tested with small probability ϵ on each time step, while Foster et al. (2000) use a soft-max selection where actions with high Q-values have a higher probability of being chosen. In robotic experiments (Cuperlier et al. 2007; Barrera and Weitzenfeld 2007), the exploration is chosen when the animat cannot associate its location with any existing node in its topological map. In Girard et al. (2005), the exploration is a random direction chosen among the other strategies, but the selection is not learned. We show here that the model in which the balance between exploitation and exploration is not predefined but learned with training can reproduce well the rat behavior in two real-world behavioral tasks. In agreement with standard RL algorithms, the exploration is mainly chosen at the beginning of the training and then decreases as the strategies are learned (Fig. 7). Our simulations also show that Planning strategy is associated with higher exploration rate (Fig. 13b, sessions 3, 6, and 9), which is explained by

the lower accuracy of the cognitive map compared to visual input (due to a limited number of nodes). In the model proposed, the path to the goal derived from the cognitive map can only follow connections between nodes, thus producing paths which are close to optimal, but still deviating from the approximately straight paths generated by Taxon strategy.

The above mentioned model suggests that exploratory behavior may be governed by a separate brain network similarly to Taxon (DLS) and planning (Hc–PFC) networks. If so, then exploratory behavior can be potentially dissociated from other strategies using a specialized experimental paradigm. In support of this idea, several experimental addressed thigmotaxic (i.e., wall-following) behavior which can be considered as an exploratory (yet non-random) behavior (Devan and White 1999; Devan et al. 1999; Pouzet et al. 2002; Chang and Gold 2004).

5.2 The mechanism of selection can result in competition and cooperation between strategies, across and within trials

In the above model, the Taxon and Planning experts learn in parallel and in such a way that action–outcome pairs generated by one of the experts can be used by the other expert to update its action value estimates. Learning of an expert from the actions performed by another expert represents cooperation between strategies in our model, which fits well the definition of cooperation introduced by behavioral studies (see Sect. 1). In our simulations, the facilitating effect of cooperation is clearly seen by observing that performance of intact simulated animals is always better than or equal to that of lesioned simulated animals, when both strategies predict correct paths (Fig. 4b, Taxon-4 and Control-4).

On the other hand, the gating network will select an expert with the highest gating value at each time step, where the gating value corresponds to the total future reward predicted for this strategy. Such a reward-based selection of experts allows competition between strategies (see Sect. 1). Evidence for competition in our simulations is given by performance data showing that when two strategies suggest contradictory predictions about goal location, lesioned simulated animals outperform control ones (Fig. 4b, Taxon-1 and Control-1 and Fig. 10b, session 10). In summary, the presented model provides a rather simple strategy selection mechanism which implements cooperation as well as competition between the strategies within the same network.

5.3 Influence of sensory cues

5.3.1 Influence of intra-maze and extra-maze cues

A noticeable contribution of the model concerns the analysis of the influence of different types of sensory cues (intra versus extramaze) on strategy selection, which is hard to do in real life experiments. Within the gating network, the gating units of both Taxon and Planning strategies receive two types of sensory input provided by Landmark Cells (i.e., landmark information) and Planning Graph nodes (location information). Essentially this means that the availability of sensory cues at each moment in time determines the relative values of available strategies. Hence, by observing the evolution of synaptic weights between sensory inputs and gating units, it is possible to assess the relative contribution of different types of input on the behavior. From the weight analysis in our simulations we make two observations.

First, in both behavioral tasks, landmark information is more important than spatial cues for strategy selection, as shown by larger average weights of Landmark Cells compared to Place Cells (bold lines versus thin lines, respectively, in Figs. 6, 11). In Simulation I, this is due to a higher accuracy of landmark information over information provided by spatial cues, since the landmark signals the correct goal location at the beginning of a session. In Simulation II, this is due to the fact that the presence or the absence of the landmark determines whether the Taxon strategy can be used at all.

Second, in Simulation II, the input from the spatial cues (i.e., Planning Graph nodes) serves mainly to decrease the influence of Taxon expert in the trials with hidden goal by negative projection from Place Cells to the Taxon gating value (Fig. 11). However, this does not completely prevent this inappropriate expert from being selected in this situation (see Fig. 13a, showing that the Taxon is selected even when it cannot "see" the landmark). In the absence of a landmark, the Taxon expert proposes a randomly chosen action and is thus equivalent to the Exploration expert. Its selection rate on the trials without landmark decreases with learning, as can be seen from Fig. 13a, b.

5.3.2 Allocentric versus egocentric cue-based learning

There are two versions of Taxon strategy in the model. They use exactly the same learning algorithm, but the visual cues are represented in an allocentric directional reference frame for the allocentric Taxon expert, and in an egocentric reference frame for the egocentric Taxon expert (Fig. 2 and Sect. 2.1). The use of allocentric directional frame implicitly requires the use of stable extra-maze cues with respect to which such a frame is defined. Our model does not include the estimation of the allocentric head direction from extra-maze cues (see Skaggs et al. 1995; Zhang 1996), but it is assumed to be provided by the head direction network (Taube et al. 1990. In contrast, infromation from the intra-maze cues is sufficient for the egocentric Taxon expert to determine direction to the goal.

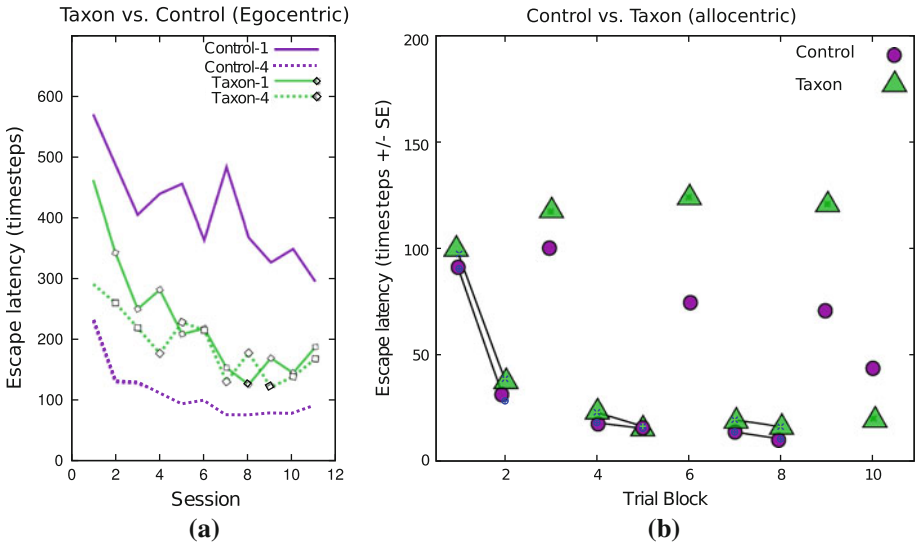

Fig. 16 **a** Results of Simulation I (Pearce et al.'s experiment) with an egocentric Taxon instead of an allocentric one. **b** Results of Simulation II (Devan and White's experiment) with an allocentric Taxon instead of an egocentric one

As shown in Fig. 9, in contrast to the egocentric Taxon strategy, the allocentric Taxon strategy reproduces the rat behavior attributed to the "heading vector" strategy observed by Pearce et al. (1998). This is because the allocentric Taxon strategy takes into account the current allocentric heading, and thus is able to tell whether the platform is located north or south of the landmark. When the platform position changes, the allocentric Taxon strategy fails to find the goal. For the egocentric Taxon strategy, the two cases are identical since the animat is using random search around the landmark in both cases.

We note here that our main results will not change if we use egocentric Taxon strategy in the simulation of the experiment of Pearce et al. (1998), as demonstrated in Figs. 4a and 16a. The use of the egocentric strategy simply slows down the performance of both Taxon and Control groups. Accordingly, the use of an allocentric Taxon strategy does not deeply change the results of Taxon and Control groups in the simulation of Devan and White (1999) when the platform is visible (Figs. 10b, 16b). However, Control group is much less efficient in hidden trials: in the sudden absence of the landmark, the allocentric Taxon, which has memorized the previous heading, helps to a lesser extent in finding the goal than does the egocentric Taxon which proposes a random orientation.

5.4 Neural substrates for the strategy-selection network

According to Ragozzino et al. (1999) and Rich and Shapiro (2009), the prelimbic–infralimbic areas (PL/IL) of the medial prefrontal cortex (mPFC) are not required for acquiring navigation strategies, but are responsible for switching between them. These data fit well to the model proposed here. Indeed, PL/IL areas receive afferents from Hc (e.g., Conde et al. 1995) and dorsomedial striatum (e.g., Groenewegen et al. 1991) which are the potential biological loci for the place- and cue-based learning, respectively. Moreover, PFC receives dopaminergic projection from the ventral tegmental area (e.g., Descarries et al. 1987). and so the reward information necessary for reward-based learning in the model may be available in the PFC.

On the neural level, Rich and Shapiro (2009) observed that different subpopulations of mPFC neurons code for different behavioral strategies. In the current model, gating values of different strategies can be considered as representing the activity of these subpopulations. Indeed, switches between strategies in the current model correspond to switch in relative gating values: if Taxon gating value is greater than Planning gating value, Taxon strategy takes the control of behavior, and vice versa (see, e.g., Figs. 6, 13). This switch between relative gating values corresponds to the switch between population activities in the recorded data of Rich and Shapiro (2009) (see Fig. 6a in their article).

Despite these similarities, however, the model cannot account for some other data in relation to the role of the mPFC in behavior. For example, it has been shown that mPFC is responsible for cross-modal but not intra-modal selection (i.e., reversal learning, Young and Shapiro 2009). In the current model, both strategy switching and reversal can be learned within the same network, since reversal in our model corresponds to simply changing the reward location. Other inconsistencies come from the study of Rich and Shapiro (2007), who have shown that mPFC is involved only during first strategy switches and it does not seem to play a role during subsequent switches. Our model cannot provide plausible explanation for these data. In summary, mPFC might be considered as a biologic locus for the selection network, but in this case (i) a separation of the gating network into at least two different parts is required to take into account the

reversal data (Young and Shapiro 2009), and (ii) an extension to the model is required to explain how the strategy switching is performed after more than a few subsequent switches (Rich and Shapiro 2007).

6 Conclusion

This study proposes a mechanism of switching between procedural cue-based and cognitive place-based navigation experts in continuous environment. The cue-based expert uses visual input, while the place-based expert uses a topological representation of the environment built on the basis of Place Cells. Random exploration is considered as a separate strategy and participates in the strategy selection process. The selection between strategies is performed by estimating how successful the strategies are in predicting the reward, on the basis of the direction of movement they propose. The model is able to select between navigation strategies that are based on distinct learning mechanisms (i.e., procedural or cognitive), potentially operating in different spatial reference frames (i.e., allocentric or egocentric). As we demonstrated, the model can serve as a useful tool for analyzing interactions between navigational strategies in spatial learning and for prediction of behaviours of lesioned animals.

The model is intended to be extended to model experimental paradigms that add, change, or remove extra-maze landmarks. The current integration of a recent hippocampal model (Ujfalussy et al. 2008) will allow Place Cells to be learned on line and to express dynamic changes in the environment. The model will also be able to simulate paradigms using multiple intra-maze landmarks. Addition of a second landmark amounts to adding another Taxon expert (either egocentric or allocentric) tuned to the new landmark. No changes need to be implemented in the selection network. Such an extended model can potentially be used to address the issue of blocking and overshadowing effects between different types of cues (Rescorla and Wagner 1972; Chamizo 2003; Gibson and Shettleworth 2003, 2005; Stahlman and Blaisdell 2009). These effects are inherent to any learning algorithm which updates associative weights between cues and rewards so as to reduce reward prediction error (e.g., TD-learning) as is true for the selection network in our model.

Acknowledgment This research was granted by the EC Integrated Project ICEA (Integrating Cognition, Emotion and Autonomy, IST 027819).

References

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

Barrera A, Weitzenfeld A (2007) Bio-inspired model of robot spatial cognition: topological place recognition and target learning. In: CIRA, pp 61–66

Blaisdell A (2009) The role of associative processes in spatial, temporal, and causal cognition. In: Watanabe SB, Blaisdell AP, Huber L, Young A (eds) Rational animals, irrational humans. Keio University Press, Tokyo pp 153–172

Brown M, Sharp P (1995) Simulation of spatial learning in the Morris water maze by a neural network model of the hippocampal formation and nucleus accumbens. Hippocampus 5(3):171–188

Burgess N (2008) Spatial cognition and the brain. Ann N Y Acad Sci 1124:77–97

Burnod Y (1991) Organizational levels of the cerebral cortex: an integrated model. Acta Biotheor 39(3–4):351–361

Canal C, Stutz S, Gold P (2005) Glucose injections into the dorsal hippocampus or dorsolateral striatum of rats prior to T-maze training: modulation of learning rates and strategy selection. Learn Mem 12(4):367–374

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

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

Chang Q, Gold PE (2004) Inactivation of dorsolateral striatum impairs acquisition of response learning in cue-deficient, but not cue-available, conditions. Behav Neurosci 118(2):383–388

Chavarriaga R, Strösslin T, Sheynikhovich D, Gerstner W (2005) A computational model of parallel navigation systems in rodents. Neuroinformatics 3(3):223–242

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 Neurorobotics 1: 1–15

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8(12):1704–1711

Descarries L, Lemay B, Doucet G, Berger B (1987) Regional and laminar density of the dopamine innervation in adult rat cerebral cortex. Neuroscience 21(3):807–824

Devan B, White N (1999) Parallel information processing in the dorsal striatum: relation to hippocampal function. J Neurosci 19(7):2789–2798

Devan B, McDonald R, White N (1999) Effects of medial and lateral caudate-putamen lesions on place-and cue-guided behaviors in the water maze: relation to thigmotaxis. Behav Brain Res 100(1–2): 5–14

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 USA 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 USA 105(15):5915–5920

Dolle L, Khamassi M, Girard B, Guillot A, Chavarriaga R (2008) Analyzing interactions between navigation strategies using a computational model of action selection. LNAI 5248:71–86

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
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
Girard B, Filliat D, Meyer J, 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
Gold P (2004) Coordination of multiple memory systems. Neurobiol Learn Mem 82(3):230–242
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. In: Willner P, Scheel-Kriiger J (eds) Functional anatomy of the ventral, limbic system-innervated striatum. 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
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
Hartley T, Burgess N (2005) Complementary memory systems: competition, cooperation and compensation. Trends Neurosci 28(4):169–170
Hasselmo ME (2005) A model of prefrontal cortical mechanisms for goal-directed behavior. J Cogn Neurosci 17(7):1115–1129
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
Kelly D, Gibson B (2007) Spatial navigation: spatial learning in real and virtual environments. Comp Cogn Behav Rev 2:111–124
Khamassi M (2007) Complementary roles of the rat prefrontal cortex and striatum in reward-based learning and shifting navigation strategies. PhD thesis, University Paris 6
Kim J, Baxter M (2001) Multiple brain-memory systems: the whole does not equal the sum of its parts. Trends Neurosci 24(6):324–330
Leising K, Blaisdell A (2009) Associative basis of landmark learning and integration in vertebrates. Comp Cogn Behav Rev 4:80–102
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):260–270
McDonald R, Devan B, Hong N (2004) Multiple memory systems: the power of interactions. Neurobiol Learn Mem 82(3):333–346
Mizumori S (2008) Hippocampal place fields. Oxford University Press, USA
O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res 34(1):171–175
O'Keefe J, Nadel L (1978) The hippocampus as a cognitive map. Oxford University Press, Oxford
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
Packard M, Hirsh R, White N (1989) Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. J Neurosci 9:1465–1472
Pearce J (2009) The 36th Sir Frederick Bartlett Lecture: an associative analysis of spatial learning. Q J Exp Psychol 62(9):1665–1684
Pearce J, Roberts A, Good M (1998) Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. Nature 396(6706):75–77
Pouzet B, Zhang W, Feldon J, Rawlins J (2002) Hippocampal lesioned rats are able to learn a spatial position using non-spatial strategies. Behav Brain Res 133(2):279–291
Ragozzino M, Detrick S, Kesner R (1999) Involvement of the prelimbic-infralimbic areas of the rodent prefontal cortex in behavioral flexibility for place and response learning. J Neurosci 19(11):4585–4594
Redish A (1999) Beyond the cognitive map: from place cells to episodic memory. The MIT Press, Cambridge
Rescorla R, Wagner A (1972) A theory of pavlovian conditioning: the effectiveness of reinforcement and non-reinforcement. In: Black A, Prokasy W (eds) Classical conditioning II: current research and theory. Appleton-Century-Crofts, New York, 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
Rich E, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. J Neurosci 29(22):7208–7219
Roberts A, Pearce J (1999) Blocking in the Morris swimming pool. J Exp Psychol Anim Behav Process 25(2):225–235
Save E, Poucet B (2000) Involvement of the hippocampus and associative parietal cortex in the use of proximal and distal landmarks for navigation. Behav Brain Res 109(2):195–206
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
Skaggs W, Knierim J, Kudrimoti H, McNaughton B (1995) A model of the neural basis of the rat's sense of direction. Adv Neural Inf Process Syst 7:173–182
Stahlman W, Blaisdell A (2009) Blocking of spatial control by landmarks in rats. Behav Processes 81(1):114–118
Strösslin T, Sheynikhovich D, Chavarriaga R, Gerstner W (2005) Robust self-localisation and navigation based on hippocampal place cells. Neural Netw 18(9):1125–1140
Sutton R, Barto A (1998) Reinforcement learning: an introduction. Bradford Book. The MIT Press, Cambridge
Taube JS, Muller RU, Ranck JBJr (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J Neurosci 10(2):420
Touretzky D, Redish A (1996) Theory of rodent navigation based on interacting representations of space. Hippocampus 6(3):247–270
Uchibe E, Doya K (2005) Reinforcement learning with multiple heterogeneous modules: a framework for developmental robot learning. In: 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
Voermans N, Petersson K, Daudey L, Weber B, Van Spaendonck K, Kremer H, Fernández G (2004) Interaction between the human

hippocampus and the caudate nucleus during route recognition. Neuron 43(3):427–435

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 (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

Yin H, Knowlton B (2004) Contributions of striatal subregions to place and response learning. Learn Mem 11(4):459–463

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

Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. J Neurosci 16(6):2112

6.5 (Tabareau et al, 2007)

Biol Cybern (2007) 97:279–292
DOI 10.1007/s00422-007-0172-2

ORIGINAL PAPER

Biological Cybernetics

Geometry of the superior colliculus mapping and efficient oculomotor computation

Nicolas Tabareau · Daniel Bennequin · Alain Berthoz · Jean-Jacques Slotine · Benoît Girard

Received: 27 October 2006 / Accepted: 4 July 2007 / Published online: 10 August 2007

Abstract Numerous brain regions encode variables using spatial distribution of activity in neuronal maps. Their specific geometry is usually explained by sensory considerations only. We provide here, for the first time, a theory involving the motor function of the superior colliculus to explain the geometry of its maps. We use six hypotheses in accordance with neurobiology to show that linear and logarithmic mappings are the only ones compatible with the generation of saccadic motor command. This mathematical proof gives a global coherence to the neurobiological studies on which it is based. Moreover, a new solution to the problem of saccades involving both colliculi is proposed. Comparative simulations show that it is more precise than the classical one.

Keywords Saccades · Superior colliculus · Spatio-temporal transformation · Computational model

This work is partly supported by the EU within the NEUROBOTICS integrated Project (The fusion of NEUROscience and roBOTICS, FP6-IST-FET-2003no. 001917).

N. Tabareau · A. Berthoz · B. Girard (✉)
UMR 7152, Laboratoire de Physiologie de la Perception et de l'Action, CNRS-Collège de France, Paris, France
e-mail: benoit.girard@college-de-france.fr

D. Bennequin
UMR 7586, Equipe Géométrie et Dynamique, Université Paris Diderot-CNRS, Paris, France

J.-J. Slotine
Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

1 Introduction

Successful goal-oriented movements rely on the ability to transform sensory inputs signaling the position of the target into appropriate motor commands. This transformation requires representation changes from the sensory input space to the motor output space. Even in the case of visually guided ocular saccades, a relatively simple sensorimotor transformation, the details of this computation are still debated.

The generation of ocular saccades greatly involves the superior colliculus (SC) (the tectum in non-mammalian vertebrates). The SC is a layered structure located in the midbrain, which receives multisensory input and accordingly generates changes in gaze orientation. It drives, in particular, the reticular formation nuclei which contain the ocular saccade motoneurons (the saccade burst generators, SBG). In the SC, the sensory inputs and the corresponding output commands are represented on retinotopic neuronal maps. Each colliculus encodes the information corresponding to the contralateral visual hemifield. A specific logarithmic deformation on the amplitude axis of this mapping was found in cats (McIlwain 1976, 1983) as well as in monkeys (Ottes et al. 1986; Robinson 1972) (see Fig. 1), whereas the mapping seems to simply be linear in the other studied species [rats (Siminoff et al. 1966), goldfish (Herrero et al. 1998), for instance]. These mappings are usually explained by a reasoning based on sensory considerations: if the projections from the retina to the SC are one-to-one and if the density of cells in the collicular maps is constant, then the absence or existence of a fovea induces linear or logarithmic mappings. We propose here an alternative approach linking these mappings with the saccadic sensorimotor transformation process.

This sensorimotor process involves the activation of a large population of cells in the motor map. This activation is centered around the position corresponding to the coordinates

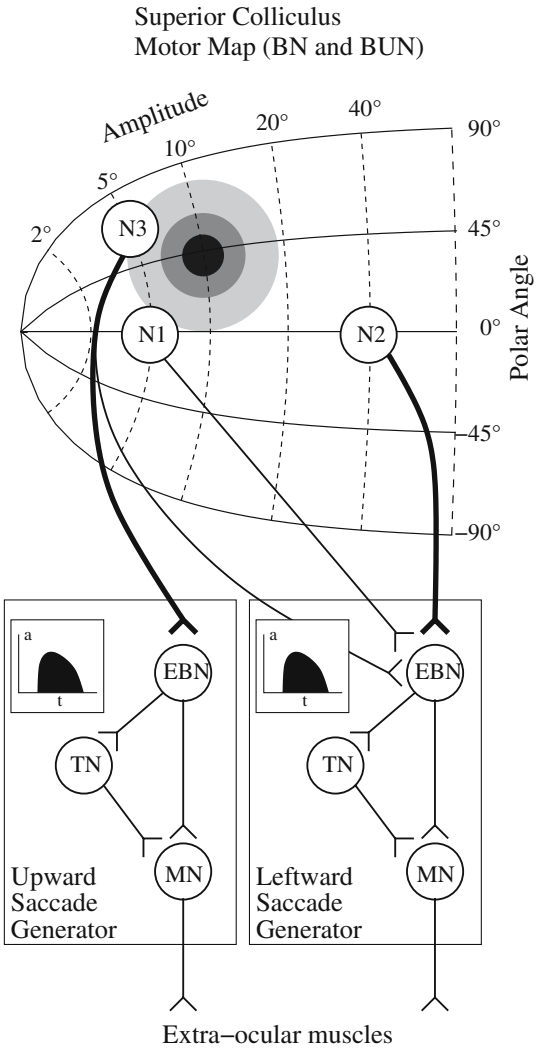

Fig. 1 Spatio-temporal transformation from the superior colliculus motor layers to the saccade generators. *BN* burst neurons; *BUN* build-up neurons; *EBN* excitatory burst neurons; *MN* motoneurons; *TN* tonic neurons. *Dashed lines* on the SC map represent iso-amplitudes and *full lines*, iso-directions. *Gray shading* on the SC map represents the activity of the population of neurons coding for a ($R = 10°, \theta = 45°$) saccade. SBG are simplified: circuitry devoted to the triggering of saccades is omitted. *Insets* represent the temporal activity of the EBNs during the execution of the saccade. The transformation from spatial to temporal coding results from selective weighted projections from SC neurons to the SBGs (strength is represented by line width): neurons $N1$ and $N2$ project to the *leftward* SBG only, as they code for horizontal saccades, and the $N2$ projection is stronger as it codes for a saccade of larger amplitude; neuron $N3$ projects to both upward and leftward SBGs as it codes for a ($R = 5°, \theta = 67.5°$) saccade

of the target of the saccade in the visual field (see upper part of Fig 1). The SBG are composed of four circuits, respectively, producing the rightward, leftward, upward and downward rotations. At this level, the movements are encoded by bursts of activity representing the vectorial components of the desired rotation (see lower part of Fig 1). The transformation from the SC distributed spatial code into the SBGs Cartesian temporal code is called the spatio-temporal transformation (STT). In addition to the problem of solving the STT for one colliculus, a *gluing problem*—in the technical sense of differential geometry (Hirsch 1976)—occurs when a vertical or quasi-vertical saccade is executed. In that case, the population activity is shared on both SC and the combination of these two activities drives the SBG. The exact location and shape of this distributed activity, and the possible role of the commissural SC projections in the coordination of the two SC, are unknown.

The first model of the STT, proposed by van Gisbergen et al. (1987), stated that it could be performed by a simple weighted sum of the activity of the SC neurons, transmitted to the SBG. This scheme has been reproduced in many early SC models [refer to (Girard and Berthoz 2005) for a review of SC and SBG models]. It assumed that the spatial shape of the activation on the SC map is stereotyped, which could be ensured by lateral connections inside the map. This model had some limitations: it did not simulate correctly the effects of simultaneous multiple site activation (saccade on the average position), of varying levels of peak activity (saccades are accurate for various peak levels of activity), and of inactivation of parts of the SC (the inactivated region "repels" saccades). The saccade averaging concern was solved in a model including lateral inhibitions within the colliculus (van Opstal and van Gisbergen 1989). However, the most important limitation is that the dynamics of appearance and disappearance of the SC activity, implying varying levels of activity, was not considered, namely, *it did not take time into account*.

In competing models, it was proposed that the output of the SC is normalized by a weighted averaging of its activity. This allowed the generation of correct saccades with varying levels of activity, and simulated the effects of multiple target averaging and of inactivation of collicular regions (Lee et al. 1988). However, as noted by Groh (2001), the division computation is critical in such a model, as it has to be carried out by a single neuron (this computation cannot be broken up among a population) and should be precise on a large range of values, which is physiologically unrealistic.

Recent experimental studies shed light on the dynamics of the saccade generation process, showing that the number of spikes produced by the whole population of SC burst neurons during saccades of different amplitudes is constant (Anderson et al. 1998; Goossens and van Opstal 2006). Moreover, it was also shown that for a given saccade, individual SC neurons always produce the same number of spikes, even in case of various kinds of perturbations: saccades interrupted by fixation zone stimulation (Munoz et al. 1996), saccades slowed by muscimol injection in omnipause neurons region (Soetedjo et al. 2000), and saccades perturbed by eye blinks (Goossens and van Opstal 2000, 2006). This strengthens recent STT models (Goossens and van Opstal 2006; Groh 2001), *which take time into account*, and where it is assumed

that an inhibitory mechanism keeps the number of spikes constant, avoiding the need for normalization. The "dynamic vector summation" model, proposed by Goossens and van Opstal (2006), implements this mechanism in a manner very similar to the Groh Groh (2001) "summation with saturation" proposal: a population of neurons sums up the number of spikes emitted by the SC and inhibits the SC output when a fixed threshold is reached. These models exhibit satisfactory behaviors in case of multiple site activation, varying levels of peak activity and inactivation of parts of the SC.

Finally, the gluing problem was addressed in the study of van Gisbergen et al. (1987). Their proposal is based on a geometrical construction which only partially uses the logarithmic mapping and systematically generates inaccurate saccades. It is also this form of gluing which was used in Goossens and van Opstal (2006).

In this work, we prove that, using a set of six hypotheses based on known neurobiology of the SC and of the SBG and fully compatible with the last two STT models (Goossens and van Opstal 2006; Groh 2001), the neural implementation of the STT is tightly linked with the geometry of the collicular mapping: it is necessarily linear or complex logarithmic. Moreover, we propose a new gluing scheme which extends these STT models to both SC, generates accurate saccades, and is compatible with the requirements of our proof.

2 Results

The quantitative description of the monkey's collicular mapping proposed by Ottes et al. (1986) can be reformulated as a complex logarithm (refer to Appendix 4.1 for more detailed considerations about quantitative description of the collicular mapping). This transformation from retinotopic Cartesian coordinates (α,β, resp. azimuth and elevation) into coordinates on the SC surface (X,Y, in millimeters) is expressed as follows:

$$\frac{X}{B_X} + i\frac{Y}{B_Y} = \ln\left(\frac{z+A}{A}\right), \quad \text{with } z = \alpha + i\beta \tag{1}$$

The values of parameters A, B_X and B_Y for the monkey have been experimentally estimated. Concerning the cat, the mapping is in accordance with such a description (McIlwain 1976), but the parameters' values have not been estimated. For animals having a linear mapping, the following formulation can be simply used:

$$\frac{X}{b_X} + i\frac{Y}{b_Y} = z \tag{2}$$

2.1 The need for a linear or complex logarithmic mapping

Our first result is a mathematical proof (detailed in Appendix 4.2) that the complex logarithmic or linear mappings (as defined by Eqs. 1, 2) are the only appropriate ones. Interestingly, these classes of mappings are conformal (as the functions are holomorphic) although it is not required by the hypotheses on which the proof is based. These hypotheses are based on the formalization of six known biological properties of the STT (their precise mathematical formulation is given in Appendix):

Weighted sum. The outputs of the SC fed to the horizontal and vertical saccade generators (SBG) are generated by weighted sums of the activity of the SC motor cells.

Glued colliculi. The two colliculi are connected with each other so that they form only one abstract mapping on the whole plane $\mathbb{R}^2$.

Invariant integral. For each motor cell, the number of spikes emitted during a whole saccade burst (without those corresponding to the eventual preceding build-up activity) depends only on its location with respect to the (X,Y) coordinates of the saccade on the collicular surface.

Linearity. The total command sent from the SC to the SBG is a linear function of z_0, the Cartesian coordinates of the saccade to be generated.

Smooth mapping. The collicular mapping is continuously differentiable. $(X, Y) = (0, 0)$ corresponds to $z = 0$, and the visual horizontal and vertical axes are aligned with the X and Y axes in 0.

Similarity. For any continuous population activity respecting the invariant integral hypothesis, the projection weights from the SC to the SBG is a similarity[1] with regards to the saccade coordinates expressed in azimuth and elevation (the retinotopic Cartesian coordinates).

This similarity hypothesis is the less intuitive of our six hypotheses as it does not seem to have any functional justification. However, if it is assumed that the mapping on each side is either linear or logarithmic (a constraint which can be due to the appearance of a fovea), we prove in Appendix 4.4 that for any activity with gaussian invariant integral, the only system of projection weights from SC to SBG (with a moderate growth), which produces a correct saccade (under the assumption of linearity), is a similarity. We also prove in this appendix that, given any activity, there is no deformation (with support on one SC) of the projection weights, except similarities, generating correct saccades. In particular this implies that the set of similarities is the only class of projection weights from SC to SBG, which can be adapted to every activity and which is stable under affine re-mapping (or modulation).

This analysis provides as a corollary an expression for the projection weights of the SC to the SBGs in the logarithmic case (it seems to be a folklore result although it never appeared in the literature). Using the equation of the logarithmic mapping, one can analytically express the projections

[1] A similarity is a transformation that preserves ratios of distances.

from the superior colliculus to the brainstem. In the special case where the coefficients a and b of the hypothesized similarity are just real numbers, we obtain:

$$\begin{aligned} w_\alpha &= aA\left(\exp\left(\frac{X}{B_X}\right)\cos\left(\frac{Y}{B_Y}\right)-1\right)+b \\ w_\beta &= aA\exp\left(\frac{X}{B_X}\right)\sin\left(\frac{Y}{B_Y}\right) \end{aligned} \tag{3}$$

A graphical representation of this analytic formulation is given in Fig. 5, upper part, using the monkey's parameters.

The fact that the total activation of one neuron on the superior colliculus during the saccade depends only on its position with regard to the point coding the saccade on the collicular surface is fundamental in inducing a logarithmic mapping. However, if we assume that the mapping is logarithmic and that inter-individual differences in the mapping parameters (A, B_X, B_Y) within one species exist, we can derive the invariant integral hypothesis from the five others (proof in Appendix 4.5).

In the course of the mathematical proof, a parameter which triggers the shape of the mapping appears: if it is null, then the mapping is linear, otherwise, it is complex logarithmic. The transition from linear to complex logarithmic is smoothly obtained by a continuous variation of this parameter. This means that during evolution, a transition from a linear to a complex logarithmic mapping could have happened without any need for changing the neural structures in charge of computing the STT.

2.2 The motor gluing of colliculi

We assumed in our *glued colliculi* hypothesis that the two colliculi are connected so that the combined activity of their motor layers can be considered as a single abstract mapping on the whole plane $\mathbb{R}^2$. To solve the gluing problem in the linear mapping case, it is sufficient to put a bump of activity in each SC at the correct position, to truncate it to keep the part within the correct visual hemifield only and then to use the sum of the activity of both colliculi to drive the SBG. However, in the case of complex logarithmic mapping, a similar approach produces systematic errors (see Sect.2.3 below).

To solve this problem, we propose another approach. It consists of progressively shifting from an activity shared by both colliculi to an activity contained by a single representation, using a modulation accounting for the closeness to the vertical axis. In this scheme, an input layer (Inp^{R} or Inp^{L}) receives activation from visual sources, independently from the activity in the contralateral visual layer (Fig. 2, upper part). These layers project to the motor layers (Mot^{R} and Mot^{L}) of the ipsilateral and contralateral colliculi (Fig. 2, lower part). The ipsilateral projections are one-to-one connections: each visual neuron projects to its homologue in the motor layer. These projections are, however, modulated by the relative part of the activity of the contralateral input layer within the boundary of its visual hemifield. This modulation is a monotone increasing function f of the subtraction of the sum of the activity within the boundary (υ^{R} and υ^{L}) to the sum of the activity in the whole map ν. The addition of a control mechanism ensuring the *invariant integral* property on the two motor maps ensures that the following holds for every saccade:

$$\begin{cases} \int_t \text{Mot}^{\text{L}}_{S^{\text{L}}_0}(S^{\text{L}},t) = \chi(S^{\text{L}}_0)\cdot\int_t \text{Inp}^{\text{L}}_{S^{\text{L}}_0}(S^{\text{L}},t) \\ \int_t \text{Mot}^{R}_{S^{\text{R}}_0}(S^{\text{R}},t) = \eta(S^{\text{R}}_0).\int_t \text{Inp}^{R}_{S^{\text{R}}_0}(S^{\text{R}},t) \end{cases} \tag{4}$$

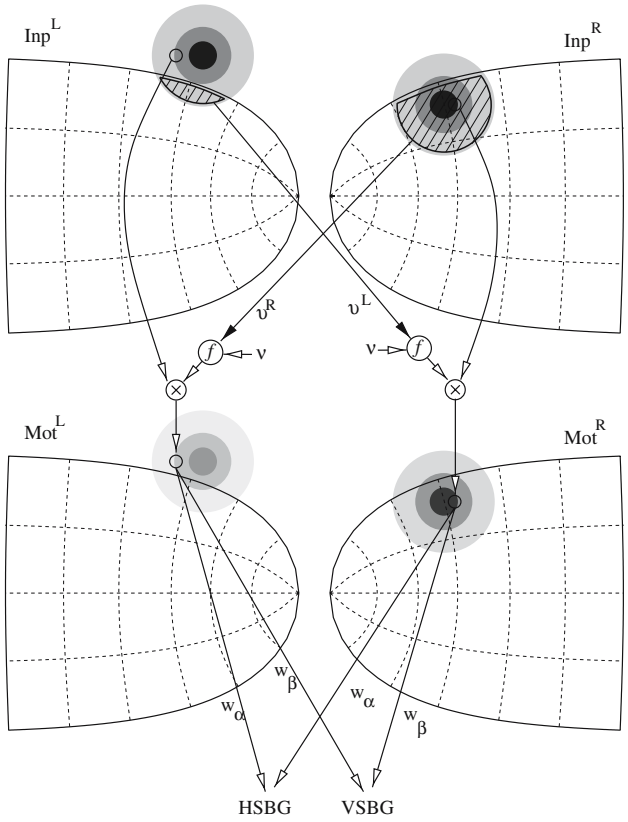

Fig. 2 Gluing method. A single target in the left hemifield but close to the vertical elicits activity in the input layers of both colliculi (Inp^{R} and Inp^{L}). In the motor layers (Mot^{R} or Mot^{L}) this activity is inversely modulated by the area of the contralateral activity within the boundary of its visual hemifield (*hatched area*, noted υ^{R} and υ^{L}). Note that ν is the sum of the activity of the *whole shaded areas*. In the motor layer, activity is, thus, much stronger in the right colliculus (coding the left hemifield) than in the left one. For a target further away from the vertical, there would be no activity left in the left motor layer. This distributed motor activity is the abstract $\mathbb{R}^2$ mapping assumed by the second property of our first proof, which can then be weighted, summed and sent to the horizontal and vertical SBG

where saccade coordinates on the left (resp. right) SC are noted S_0^L (resp. S_0^R). The positive functions χ, η are the result of the integrated commissural modulation and satisfy $\chi(S_0^L) + \eta(S_0^R) = 1$ for all saccade. This constraint ensures that the sum of the activity on both colliculi behaves exactly as a single activity on an abstract map (a complete description of the scheme is given in Appendix).

2.3 Simulation

To assess the accuracy of this gluing scheme, and also to compare it with the proposal of van Gisbergen et al. (1987), we built a simple computational model of the SC and SBG based on the Groh architecture for STT (Groh 2001) (see Fig. 7). This model is made of rate-coding leaky-integrator neurons. Each SC contains two 90 × 90 neuron maps, a visual input one and a motor one, respecting the monkey mapping equation from Ottes et al. (1986). The activity generated by a target is a 2D Gaussian ($\sigma = 0.5$ mm) centered on the target coordinates expressed in the collicular mapping. The activity of the motor map is controlled by a *summation with saturation* architecture. The SBGs' implementation is minimal, they contain no feedback loop, and are made of inhibitory and excitatory burst neurons receiving the output of the SC motor layer, of tonic neurons integrating the burst neurons activity and of motoneurons summing the burst and tonic neuron outputs. The eye plant is simulated by the standard second-order differential equation model, linking eye rotation and the motoneuron firing rate. Details of the model are given in Appendix 4.7.

In the van Gisbergen et al. (1987) proposal for gluing, when a saccade is so close to the vertical that the activity on the SC crosses the 90° or −90° iso-direction curves, a second bump of activity is placed in the other SC, and the two bumps are truncated to keep the part within the preferred hemifield of each SC only. However, rather than using the mapping Eq. (1), they use a *ad-hoc* geometrical construction to place the second bump. This construction generates systematic errors for saccades close to the vertical (see their Fig. 4). In our simulation, we tested their truncation gluing scheme, but positioned the second bump according to the mapping equations. Even with this enhancement, relatively large systematic saccades errors are generated: the upper part of Fig. 3 shows the error (measured as the distance between desired and effective saccade endpoint divided by the desired saccade amplitude) for saccades generated over the [0°, 14°] horizontal interval and the [10°, −10°] vertical interval, with a 1° increment. This error, which reaches more than 27% for the (1°, 0°) saccade, is around 5% in the vertical region, where gluing occurs.

The implementation of our model produces errors that are comparatively much lower (less than 1.5%, lower part of Fig. 3). These errors are caused by integration approximations when numerically solving the model's differential equations and by the coarse discretization of the SC, rather than by an approximate gluing.

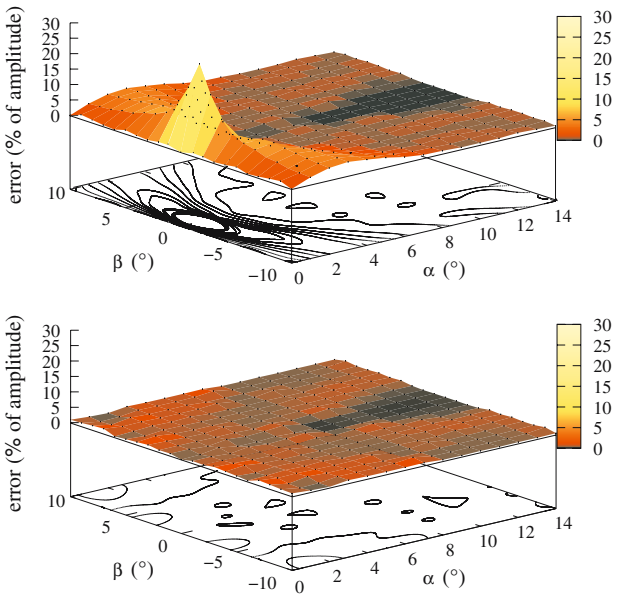

Fig. 3 Saccades endpoint error maps. 3D representation of the ratio of the distance between desired and generated saccade endpoints and the amplitude of the desired saccade, for saccades generated by the van Gisbergen et al. gluing scheme (*top*) and our proposal (*bottom*). Note that the van Gisbergen et al. proposal generates systematic errors close to the vertical, the result of an incorrect gluing. α azimuth; β elevation

3 Discussion

We showed that collicular mapping has to be either linear or logarithmic in order to control the SBG correctly, assuming six basic properties of the spatio-temporal transformation. This result also shows that a continuous transition from the linear to the logarithmic mapping can be made, affecting neither the neural substrate nor the underlying computations generating saccadic movements. In an evolutionary perspective, it suggests that the appearance of a fovea and the corresponding modification of the mapping of the visual areas could have happened in a progressive manner without requiring any modification of the final stages of the saccadic circuitry.

A hypothesis of this first result is that the two colliculi have to be combined so as to be equivalent to a single abstract mapping of the whole visual field. We, thus, proposed a new gluing scheme which generates saccades of the correct size and predicts the role and structure of the commissural projections in charge of driving this motor gluing.

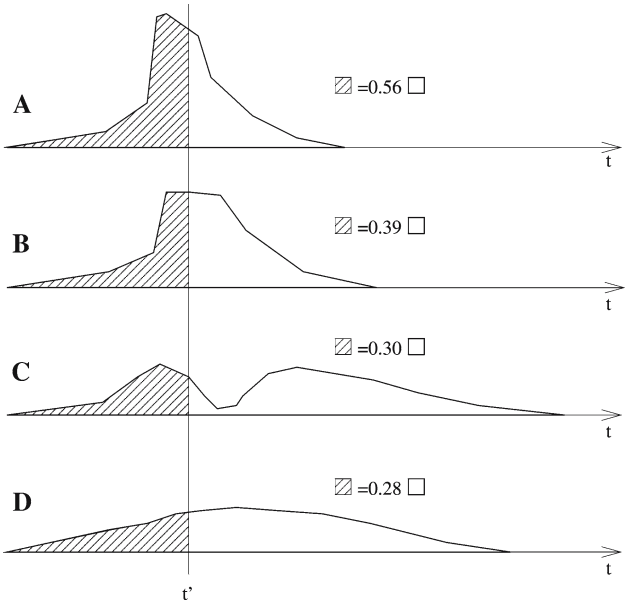

Fig. 4 Consequence of the invariant integral hypothesis for one SC neuron. *A, B, C* and *D* are schematic drawings of the activity of a given neuron of the SC, for a given saccade metric. While A represents a normal saccade. B, C and D represent the activity of the neuron in perturbed saccades (like during stimulation of the fixation cells, muscimol injection in the OPNs, eye blink, etc.). In all cases, the integrated activity over the whole burst duration (the surface within the *bold polygon*) is constant; thus, these activations are compatible with the invariant integral hypothesis. Note that, at a given moment t', the generated fraction of this activity (represented by the *hatched surface*) may vary

3.1 The six basic properties

We first discuss the neurobiological relevance of the six properties on which we based our proof, for the monkey and the cat.

The *weighted sum* property corresponds to the simplest way to transmit the activity of a population of SC neurons to the SBG, as no additional circuitry is needed between SC motor cells and SBG bursters in order to, for example, select the most active neuron only. Moreover, relying on such a population coding is more resilient to noise in neural activity. This hypothesis has received support from both experimental (Moschovakis et al. 1998; Sparks et al. 1976) and modeling (Badler and Keller 2002; van Gisbergen et al. 1987) studies.

The *invariant integral* property states that the shape of the activity on the SC map $\mathcal{A}$ does not have to be perfectly invariant in space and time, as long as the activity of each cell integrated over saccadic signal duration (i.e., number of spikes emitted during the saccadic burst) depends only on its location with regards to the point on the SC surface coding for the saccade metrics. This hypothesis is weaker than the invariant Gaussian used in numerous models, it avoids putting too much constraint on the precise tuning of the activity profiles of the SC neurons (as depicted in Fig. 4). Not demanding temporal stereotypy allows the duration of a saccade of a given metric to vary from one execution to another, for example, because of varying peak levels of activity, as long as the integrated activity is constant.

As mentioned in the Introduction, a number of recent experimental studies (Anderson et al. 1998; Goossens and van Opstal 2000, 2006; Munoz et al. 1996; Soetedjo et al. 2000) with monkeys show that the number of spikes emitted by a given SC neuron for a given saccade is constant, despite various types of perturbation. This fully supports our *invariant integral* hypothesis, at least for monkeys. We are not aware of similar results in cats that could shed a complementary light on our *invariant integral* hypothesis. Since the morphology and physiology of SC neurons is quite different in felines and primates (Grantyn and Moschovakis 2003), such studies are necessary to test the validity of the hypothesis in cats.

The *linearity* property states that the desired saccade amplitude has to linearly depend on the SC output. The burst neurons of the SBG, which receive this SC output and generate the phasic part of the motoneuron activity responsible for saccadic eye movement, exhibit an affine relationship between the number of spike they emit during a saccade and the amplitude of the saccade, in monkeys (Keller 1974; King and Fuchs 1979) as well as in cats (Kaneko et al. 1981; Yoshida et al. 1982). If the summed offsets of the affine functions of the burst neurons coding for two opposite directions are equal, then the fact that the SBG are controlled linearly holds. It happens that the SBG also receives input from the fastigial oculomotor region (FOR) of the deep cerebellar nuclei. It does not affect our proof, as we do not demand that the SBG input exclusively comes from the SC. However, it means that this affine relationship in the burst neurons is not the result of the SC influence only. Thus, the SC input signal might vary non-linearly with saccade amplitude, as the cerebellar input could compensate this non-linearity, so that the summed command remains linear. However, it was shown (Iwamoto and Yoshida 2002) that in monkeys, an inactivation of FOR results in a saccadic gain modification. This means that the suppression of the FOR input to the SBG generates saccadic movement whose amplitude still varies linearly with the amplitude of the desired saccade, proving that the collicular input to the SBG is also a linear command, whose gain is not 1, and that has to be compensated by cerebellar input. Concerning cats, the effects are affecting either the gain or the offset for, respectively, contraversive and ipsiversive movements (Goffart 1998). However, this study was carried out head-free, similar head-fixed experiments would be necessary to validate or invalidate our hypothesis in cats.

The *similarity* property states that the projection weights from the SC to the SBG are a similarity of the saccade coordinates, *expressed in the visual space*. This unintuitive

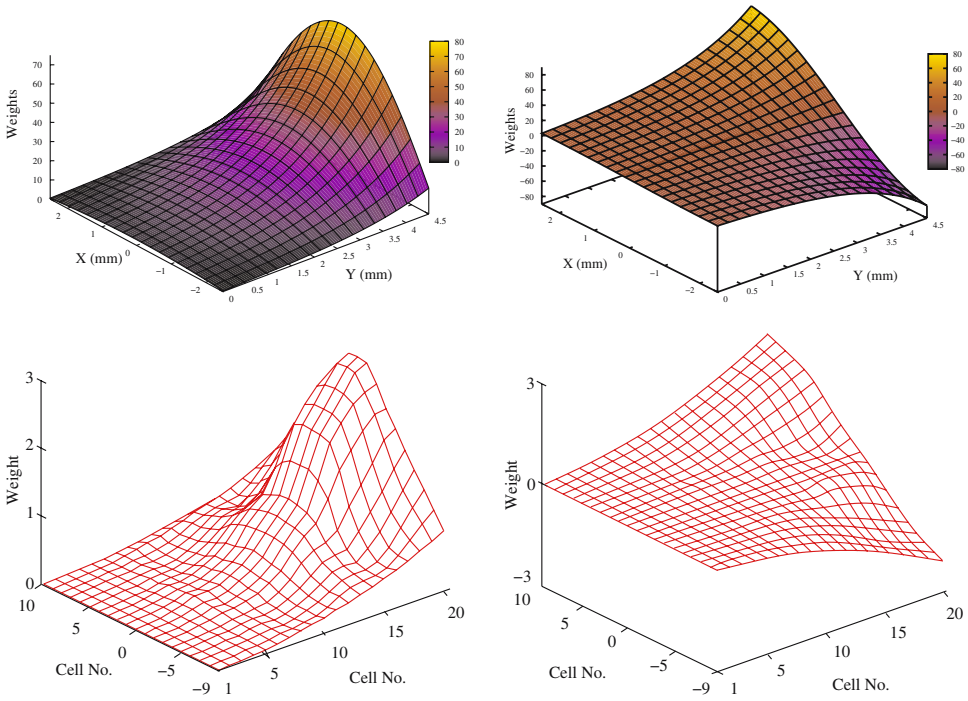

Fig. 5 *Top* plots of the weight from all the motor cells of the superior colliculus to the brainstem horizontal burst generator (*left*) and to the vertical burst generator (*right*) in monkeys. The values of these weights were obtained by the exact equations described in text Eq. (3) and parameters specific to the monkey ($A = 3°$, $B_X = 1.4$ mm and $B_Y = 1.8$ mm and $a = 1$, $b = 0$). *Bottom* plots of the weights obtained by Arai et al. (1994) with a learning algorithm for a map covering from 0° to 20° in amplitude and from −65° to 65° in direction

property was indeed derived from the evidence that in cats this projection is affine on the horizontal axis (Moschovakis et al. 1998). However, neither the fact that the vertical projection is affine nor the fact that the whole projection function is a similarity, a subset of the affine functions, were proved in cats. Moreover, no result of that type is available for monkeys. However, as evoked in the Results section, using the five other hypotheses and assuming that the mapping is either linear or complex logarithmic, we were able to prove that the weights respect the similarity hypothesis.

The Appendix 4.6 of this paper contains a generalization of our results, showing that if we relax the *similarity* hypothesis by assuming affine projection only, three additional types of mappings become acceptable and all the resulting five mappings can be non-linearly twisted. Finding animals whose mapping corresponds to one of these three mappings would favor the affine hypothesis.

Note that this hypothesis is formulated so that *similarity* has to be true for any $\mathcal{A}$. Thus our result implies that with a complex logarithmic mapping, for any $\mathcal{A}$ function verifying the *invariant integral* property, the parameters a and b (defining the weights in Eq. 3) can be found so that a weighted sum of the activity of the SC neurons will generate accurate saccades. This means that the precise shape of $\mathcal{A}$ can change during lifetime and be different from one individual to another: an adaptive mechanism tuning a and b is sufficient to ensure correct operation of the system, there is no need for changing the mapping of the SC maps itself.

The hypothesis, that the *similarity* must be true for any $\mathcal{A}$ function verifying the property of *invariant integral*, is quite strong. However, our proof holds true even with restricted families of activations. For example, if the *similarity* has to be true for Gaussian functions with small perturbations of mean, we still obtain the two mappings.

Concerning the *smooth mapping* property, stating that the mapping function ϕ is continuous comes directly from the well known retinotopy of SC maps. Stating that its first derivative is also continuous means that the variation of the magnification factors on the maps are smooth, which is verified in all studied species. Finally, the X and Y axes used to describe the maps are chosen, by convention, so as to be aligned with the horizontal and vertical directions in 0.

The fact that these neurobiological properties and the known SC mappings can be combined together in a mathematical proof strengthens their coherence and reduces the concern of their individual uncertainties. Experimentally exploring the validity of these six properties in species other than cat and monkey, especially those having a linear mapping, could reveal whether our results can be generalized among vertebrates.

3.2 SC to SBG projection weights

As regards the projection from the superior colliculus to the saccade generator, we must say that to our delight similar profiles have been obtained by Arai et al. (1994) using a training

procedure based on their model of the SC (see Fig. 5). It shows both that these weights can be obtained by learning and that our theoretical approach is corroborated by a more experimental one. Nevertheless, in another paper (Arai et al. 1999), they obtained different profiles as they used a mixed velocity and position feedback to control SC activity, which transgresses our *invariant integral* hypothesis.

A few neurobiological studies tried to evaluate the weights of the connections from the SC to the SBG. The density of SC neurons projecting to the horizontal SBG in monkeys (Grantyn et al. 2002) have variation tendencies compatible with our results, at least for a range of saccades for which head movement are negligible. The technology available to estimate projection weights is however, too limited yet to provide a full account of or to reject our result.

3.3 Is there a STT?

Optican (2005) proposes that the sensorimotor transformation necessary to convert visual input into motor command does not need to be explicitly performed as a STT between the SC and the SBG. In his model, the SC gives only an initial directional drive to the saccadic system, while the cerebellum plays the major part, as it implicitly performs the transformation.

It can be reasonably assumed that the importance of the cerebellum has been neglected in previous modeling studies, as its role in the calibration of the system and in on-line adjustments of saccade trajectory is fundamental. It could indeed replace the reticular formation displacement integrator postulated by many former SBG models. Nevertheless, the available neurobiological data, that we use to build our proof, clearly shows that all the elements needed to perform a STT between the SCs and SBGs are present. We thus propose that a STT indeed occurs, with a gain different from 1, and that the cerebellum constantly compensates for this difference.

3.4 Commissural projections

Commissural projections seem to exist at every level of the SC (Olivier et al. 1998), and many of them are probably used to solve various gluing problems, such as ensuring consistency of visual information in the superficial layers, or continuity of retinotopic working memory at the level of the quasi-visual cells. Our proposal uses a set of commissural projections to solve the gluing problem at the motor level, and thus makes predictions concerning these commissural projections only. Experimentally distinguishing these various types of commissural projection might be crucial for the understanding of their organization and roles.

4 Appendix

4.1 Coordinates on the SC layers and mapping formulation

The question of the nature of the coordinate system that should be used to describe the mapping on the collicular layers has to be raised. Indeed, the colliculus, and especially its superficial visual layers, are convex. The maps proposed in biological studies are obtained with various methods: projections on the Horsley–Clarke plane (Dräger and Hugel 1976; Feldon et al. 1970; Robinson 1972; Siminoff et al. 1966), empirical flattening of the surface by cutting (Rosa and Schmid 1994), or locally cylindrical coordinates (Knudsen 1982). None of these methods respects the curvature of the surface. Only Siminoff et al. (1966) propose a correction—on two axes only rather than for the whole surface—that takes the curvature into account.

Solving this question is beyond the scope of this paper, we however stress that our results concern the activity of the intermediate motor layers of the colliculus, which seems to be much more planar, or at least unfoldable. We will, therefore, use a Cartesian coordinate system (X, Y) to localize points on the surface of these intermediate or deep layers.

Two-dimensional saccades result from the conjunction of the activity of horizontal and vertical brainstem generators. So the final motor coordinate system is a priori a Cartesian one. However, Robinson (1972) has shown that for the monkey, the sensorimotor maps of the SC are more adequately described by a deformed polar coordinate system.

The equations mapping retinotopic polar coordinates (R,θ) onto the collicular surface (Cartesian coordinate (X,Y) in millimeter), first introduced by (Ottes et al. 1986), are

$$X = B_X \ \ln\left(\frac{\sqrt{R^2 + 2AR\cos(\theta) + A^2}}{A}\right) \tag{5}$$

$$Y = B_Y \ \text{atan}\left(\frac{R\sin(\theta)}{R\cos(\theta) + A}\right) \tag{6}$$

With the following parameter settings: $A = 3.0°$, $B_X = 1.4\,\text{mm}$ and $B_Y = 1.8\,\text{mm}$. Even if a precise evaluation of these parameters for the cat was not provided, the cat's mapping depicted in (McIlwain 1976) seems to be in accordance with such a description, with a B_Y/B_X ratio close to 2.

As noted in (Ottes et al. 1986), this mapping can however be reformulated as complex logarithm of a linear function of eccentricity, as proposed by Schwarz (1980) in its modeling of the striate cortex mapping. Using z, the complex variable defined as:

$$z = \alpha + i\beta \tag{7}$$

where α and β represent the horizontal and vertical amplitude of the saccade, one can rewrite Eqs. 5 and 6:

$$\frac{X}{B_X} + i\frac{Y}{B_Y} = \ln\left(\frac{z+A}{A}\right) \tag{8}$$

4.2 The need for a linear or complex logarithmic mapping

The keystone of our result lies in a mathematical formulation of the six biological properties of the spatio-temporal transformation as equations in $\mathbb{C}$.

We work with a complex formulation $S = X + iY$ of coordinates on the abstract SC map together with a bijection ($z = \phi(S)$) from the colliculus map to the visual hemifield. All along the proof, we will refer to a given desired saccade $z_0 = \alpha_0 + i\beta_0$ in visual coordinates which can be expressed in collicular coordinates as a specific $S_0 = \phi^{-1}(z_0)$. The command sent from the superior colliculus to the saccade generators (H: horizontal, V: vertical) in order to generate a given S_0 saccade is described by $\mathrm{Out}_{S_0}(t) = \mathrm{Out}^{\mathrm{H}}_{S_0}(t) + i\,\mathrm{Out}^{\mathrm{V}}_{S_0}(t)$. We can now formulate the weighted sum property as

$$\mathrm{Out}_{S_0}(t) = \int_S w_S \mathcal{A}_{S_0}(S,t)\mathrm{d}S \tag{9}$$

where $w_S \in \mathbb{C}$ are the weights of connection from the neuron located in S to the saccade generators and $\mathcal{A}_{S_0}(S,t) \in \mathbb{R}$ is the activity on the abstract map at location S and time t for a S_0 saccade. Technically, $\mathcal{A}_{S_0}$ is a function such that for all fixed t, the product of $\mathcal{A}_{S_0}(\_,t)$ with any exponential function is of finite integral. For example, it can be a Gaussian or any function with compact support (which will be the case in practice). Similarly, the invariant integral property amounts to say that there exists a function $K_{\mathcal{A}}$ such that

$$\int_t \mathcal{A}_{S_0}(S,t)\mathrm{d}t = K_{\mathcal{A}}(S - S_0) \tag{10}$$

and the linearity property expresses that

$$\int_t \mathrm{Out}_{S_0}(t)\mathrm{d}t = Cz_0 \quad (C \in \mathbb{R}) \tag{11}$$

The similarity property states that for any activation $\mathcal{A}$ that satisfies (10), w_S is a similitude in z. This is equivalent to the existence of two complex numbers a and b such that

$$w_S = az + b \tag{12}$$

Asking for a smooth mapping means that $\phi \in \mathcal{C}^1$, satisfies $\phi(0) = 0$, and is aligned with the X and Y axes in 0 ($D_X\phi(0) \in \mathbb{R}^+$ and $D_Y\phi(0) \in i\mathbb{R}^+$).

From Eqs (9), (10) and (11), it is easy to derive:

$$C\phi(S_0) = \int_S w_S\, K_{\mathcal{A}}(S - S_0)\mathrm{d}S \tag{13}$$

We will differentiate this equation with respect to X and Y. Let ψ be either $D_X\phi$ or $D_Y\phi$ and $K = K_{\mathcal{A}}$. Using Eq. (12) and the fact that $z = \phi(S)$, we get

$$\forall S_0 \quad C\psi(S_0) = a\int_S K(S - S_0)\psi(S)\mathrm{d}S \tag{14}$$

note that C, a and K depend on $\mathcal{A}$ which is not the case for ψ. We now use the possibility to choose different functions for the activity and translate the activity $\mathcal{A}$ for small vectors u. We pose $\kappa(\mathcal{A}) = C/a$ and introduce the notation $f_u(S) = f(S+u)$ for any function f

$$\begin{aligned}\kappa(\mathcal{A}_u)\psi(S_0) &= \int_S K(S - (u + S_0))\psi(S)\mathrm{d}S \\ &= \int_S K(S' - S_0)\psi(S' + u)\mathrm{d}S \\ &= \kappa(\mathcal{A})\psi_u(S_0)\end{aligned}$$

Let $F(u) = \kappa(\mathcal{A}_u)/\kappa(\mathcal{A})$. We have for any small u

$$\psi(S+u) = \psi(S)F(u)$$

Applying this to $S = 0$ leads to $F(u) = \psi(u)/\psi(0)$, so

$$\psi(S)\psi(u) = \psi(S+u)\psi(0) \tag{15}$$

Let us introduce the change of coordinates $S \mapsto \widetilde{S}$ that makes the Jacobian of $\widetilde{\phi}$ (i.e., the function ϕ in the new coordinates) equal to I at 0. By the hypothesis of smooth mapping, we know that

$$\widetilde{S} = \frac{X}{b_X} + i\frac{Y}{b_Y} = \widetilde{X} + i\widetilde{Y} \tag{16}$$

for some $b_X, b_Y \in \mathbb{R}^+$. By the (Theorem 1, p. 225 of Bourbaki 1972), we know that ψ is analytic. We then deduce from Proposition 7, p. 200 of the same book that ψ is an exponential function, i.e.,

$$\exists\, C_1, C_2, \lambda, \mu \text{ in } \mathbb{C} \quad \begin{cases} D_X\widetilde{\phi}(\widetilde{S}) = C_1\exp(\lambda\widetilde{X} + \mu\widetilde{Y}) \\ D_Y\widetilde{\phi}(\widetilde{S}) = C_2\exp(\lambda\widetilde{X} + \mu\widetilde{Y}) \end{cases} \tag{17}$$

Applying Schwarz's theorem which states that the partial derivatives commute, we infer that $\mu = i\lambda$. To integrate those equalities and obtain the different forms of $\widetilde{\phi}$, we have to distinguish between two cases.

- $\lambda \neq 0$
  in that case,

$$\frac{1}{\lambda}(\exp(\lambda\,\widetilde{S}) - 1) = z, \quad \lambda \in \mathbb{C} \tag{18}$$

which can be rewritten if $\lambda \in \mathbb{R}^+$

$$\frac{X}{B_X} + i\frac{Y}{B_Y} = \ln\left(\frac{z+A}{A}\right) \tag{19}$$

- $\lambda = 0$
  in that case,

$$\widetilde{S} = z \tag{20}$$

which can be rewritten

$$\frac{X}{b_X} + i\frac{Y}{b_Y} = z \tag{21}$$

Remark that this case is simply the limit case of the exponential mapping when $\lambda \to 0$.

For our proof to be complete, it remains to check that the necessary conditions found above are also sufficient by explicitly computing a and b. To make the formulations simpler, we introduce $u = S - S_0$ and $\widetilde{u} = \widetilde{S} - \widetilde{S}_0$

$$\begin{cases} a = C\left(\sum_u \exp(\lambda\,\widetilde{u})\cdot K_{\mathcal{A}}(u)\right)^{-1} \\ b = \frac{C}{\lambda}\left(\left(\sum_u \exp(\lambda\,\widetilde{u})\cdot K_{\mathcal{A}}(u)\right)^{-1} - \left(\sum_u K_{\mathcal{A}}(u)\right)^{-1}\right) \end{cases} \tag{22}$$

4.3 The gluing of the two colliculi

In order to satisfy the *glued colliculi* hypothesis of the above proof, we propose a method for gluing the colliculi so that we can then consider them as a single abstract mapping ϕ on the whole plane.

We define two distinguished layers Inp and Mot and connect them by direct and commissural connections, as depicted in Fig. 2. The SC neurons sending commissural projections are confined within the boundary of the preferred hemifield, defined by the iso-direction curves 90° and −90° (hatched areas in Fig. 2). By defining the T operator as 1 within this boundary and 0 outside, we can mathematically express υ^{L} and υ^{R} as follows:

$$\begin{aligned} \upsilon^{\mathrm{L}} &= \sum_S T(S)\mathrm{Inp}^{\mathrm{L}}(S,t) \\ \upsilon^{\mathrm{R}} &= \sum_S T(S)\mathrm{Inp}^{\mathrm{R}}(S,t) \end{aligned} \tag{23}$$

The sum of the whole activity in one input layer, ν, is defined as:

$$\nu = \sum_S \mathrm{Inp}^{\mathrm{R}}(S,t) = \sum_S \mathrm{Inp}^{\mathrm{L}}(S,t) \tag{24}$$

We can then relate the four layers by the following equations:

$$\begin{aligned} \mathrm{Mot}^{\mathrm{L}}(S^{\mathrm{L}},t) &= f(\upsilon^{\mathrm{R}})\mathrm{Inp}^{\mathrm{L}}(S^{\mathrm{L}},t) \\ \mathrm{Mot}^{\mathrm{R}}(S^{\mathrm{R}},t) &= f(\upsilon^{\mathrm{L}})\mathrm{Inp}^{\mathrm{R}}(S^{\mathrm{R}},t) \end{aligned} \tag{25}$$

where S_0^{L} and S_0^{R} are the saccade coordinates expressed in the left and right collicular mappings, S^{L} and S^{R} the coordinates of the considered neuron in the left or right SC, and f is a transfer function tuned to be highly receptive when half of the activity bump enters the boundary of the preferred hemifield. For that, f is a sigmoid with a high steepness ρ, centered at one half of ν:

$$f(x) = 1 - \frac{1}{1+\exp^{\rho(0.5\nu - x)}}$$

The invariant integral property ensures that this four layered structure satisfies relation (4) with $\chi + \eta = 1$.

We will call abstract map the result of this gluing.

4.4 Proof of the stability of the similarity under small deformations

We suppose in this paragraph that the collicular mapping is either logarithmic or linear, and that the collicular output is linear (as required by the linearity hypothesis).

We choose the complex coordinates $S = X + iY$ on the colliculus and z in the visual field, such that for any S, $z = \phi(S) = \exp(S) - 1$ or $z = \phi(S) = S$. We have seen that for any given kernel $K_{\mathcal{A}}$ resulting from an invariant integral activity $\mathcal{A}$, integrable with every exponential weights, there exists a similarity σ in the z-plane,such that for any S_0:

$$\phi(S_0) = \int_S K_{\mathcal{A}}(S - S_0)\sigma(\phi(S))\mathrm{d}S. \tag{26}$$

Suppose that $K_{\mathcal{A}}$ is non identically zero; we want to prove that σ is the only function satisfying this equation under natural growth conditions.

Recall the Fourier–Laplace transform of a function (or a distribution) u in the S-plane is defined in a point $\zeta = (\xi, \eta)$ in $\mathbb{C}^2$, by the integral:

$$\widehat{u}(\zeta) = \int u(X,Y)\mathrm{e}^{-i(\xi X + \eta Y)}\mathrm{d}S, \tag{27}$$

when this integral converges.

We put the hypothesis on $K_{\mathcal{A}}$ that its Fourier–Laplace transform $\widehat{K}$ is defined and complex analytic over the entire complex plane.

Let $f : \mathbb{R}^2 \to \mathbb{C}$ be a continuous function, satisfying the equation (26):

$$\phi(S_0) = \int_S K_{\mathcal{A}}(S - S_0)f(S)\mathrm{d}S. \tag{28}$$

Our hypothesis will be that there exists a constant c and and an open set Ω in $\mathbb{C}^2$ containing a plane parallel to $\mathbb{R}^2$, such that the difference $\Delta = f - \sigma \circ \phi - c$ has a well defined Fourier–Laplace transform on Ω. This is verified if the gradient $\nabla \cdot \Delta$ is equal to zero for X sufficiently negative or $|Y|$ sufficiently large and is majored by an exponential function for X positive. We remark that the preceding condition means that the deformation Δ is supported by one of the two colliculi.

Let us denote by $D\Delta$ either $D_X\Delta$ or $D_Y\Delta$; from equation(26), we have:

$$\int_S K_{\mathcal{A}}(S - S_0) D\Delta(S) \mathrm{d}S = 0. \tag{29}$$

Thus (Hörmander 1983) for ξ in Ω we obtain

$$\widehat{K}(\xi)\widehat{D\Delta}(\xi) = 0. \tag{30}$$

But when the product of two analytic functions is zero, one of the functions is identical to zero. As $\widehat{K}$ is not identical to zero, $\widehat{D\Delta}$ is zero on Ω, and by the injectivity of the Fourier–Laplace transform (Hörmander 1983), $D\Delta$ itself is zero. Thus f is a similarity. □

Remark 1. If $K_{\mathcal{A}}$ and $D\Delta$ were continuous functions (or even distributions) with compact support (which is not so restrictive when considering neural activity on SC maps), we could have directly deduced the result $D\Delta = 0$ from the classical "Theorem of Supports" of Titchmarsh and Lions Hörmander (1983).

2. It is not true in general that equation (26) has a unique solution, for example if the total integral of K is zero we can add any constant to σ; moreover if the Fourier transform of K becomes zero at some points in $\mathbb{R}^2$ there exists non-trivial polynomial function Δ verifying (29), their Fourier transform having support reduced to isolated points. This phenomenon cannot appear when K belongs to the class of " Wiener functions", which by definition have Fourier transforms without zero, in this case Δ can be any tempered distribution in the sense of Schwartz and we deduce $D\Delta = 0$.

In the special case of Gaussian integral of activities $K_{\mathcal{A}}$, the restrictive hypothesis on Δ can be greatly weakened: we only have to require that there exists two real constants α, β such that Δ has a well defined Fourier–Laplace transform, as an element of the space $\mathcal{S}'$ of Schwartz tempered distributions, on the plane $\Pi = (i\alpha + \xi, i\beta + \eta) | (\xi, \eta) \in \mathbb{R}^2$ in $\mathbb{C}^2$. For example Δ can be any function with polynomial growth times an exponential. Let us prove that this condition is sufficient to imply $D\Delta = 0$:

By hypothesis there exists a positive symmetric two by two matrix of determinant one A, a point M_O in $\mathbb{R}^2$, a constant $C > 0$ and a number $\tau > 0$, such that

$$K(S) = C(4\pi\tau)^{-1} \mathrm{e}^{-(S-M_0)\cdot A(S-M_0)/4\tau}. \tag{31}$$

The Fourier–Laplace transform of K is the analytic function

$$\widehat{K}(\zeta) = C\mathrm{e}^{-\tau\zeta\cdot A^{-1}(\zeta) + i\zeta\cdot M_0} \tag{32}$$

On the other hand $D\Delta$ has a Fourier–Laplace transform, well defined as a tempered distribution on the plane Π, it is the product of $\widehat{\Delta}$ with the restriction to Π of a linear form on $\mathbb{C}^2$. The convolution equation satisfied by $D\Delta$ implies as before $\widehat{K}\widehat{D\Delta} = 0$, but $\widehat{K}$ restricted to Π belongs to the test space $\mathcal{S}$ of Schwartz functions with quick decreasing at infinity, and has no zero at all, so $\widehat{D\Delta} = 0$. □

4.5 Proof of the need for an invariant integral

We now replace the invariant integral property by the fact that any logarithmic mapping works. We show that Eq. 10 can be deduced, i.e.,

$$\int_t \mathcal{A}_{S_0}(S, t)\mathrm{d}t = K_{\mathcal{A}}(S, S_0) = K_{\mathcal{A}}(S - S_0, 0) \tag{33}$$

Using Eqs. 9, 11 and 12 for any logarithmic mapping leads to

$$\forall\lambda \quad Ce^{\lambda S_0} = \int_S (a\mathrm{e}^{\lambda S} + b) K_{\mathcal{A}}(S, S_0)\mathrm{d}S \tag{34}$$

We differentiate with respect to λ

$$\forall\lambda \quad \kappa(\mathcal{A})e^{\lambda S_0} = \int_S e^{\lambda S} K_{\mathcal{A}}(S, S_0) dS \tag{35}$$

We pose $\Delta(S, S_0) = K_{\mathcal{A}}(S, S_0) - K_{\mathcal{A}}(S - S_0, 0)$ and deduce that

$$\forall\lambda \quad \int_S \mathrm{e}^{\lambda(S-S_0)} \Delta(S, S_0)\mathrm{d}S = 0 \tag{36}$$

The Laplace transform is defined for $\mathcal{A}$, thus, it is also defined for $K_{\mathcal{A}} = \int_t \mathcal{A}\mathrm{d}t$, as the integration is on a finite interval, and finally for any difference of to such K functions, in particular for Δ. Then, with the same argument as in Sect. 4.4, we have that

$$\Delta(S, S_0) = 0 \tag{37}$$

4.6 A generalization of the similarity hypothesis

If we relax the hypothesis of similarity by just asking for an affine projection, i.e., $w_S = \mathbf{a}z + b$ where $\mathbf{a}$ is a 2×2 invertible matrix, we then get five types of solutions.

Indeed, denoting by $J(S)$ the Jacobian of ϕ at point S leads to

$$J(u)J(0)^{-1}J(S_0) = J(u + S_0) \tag{38}$$

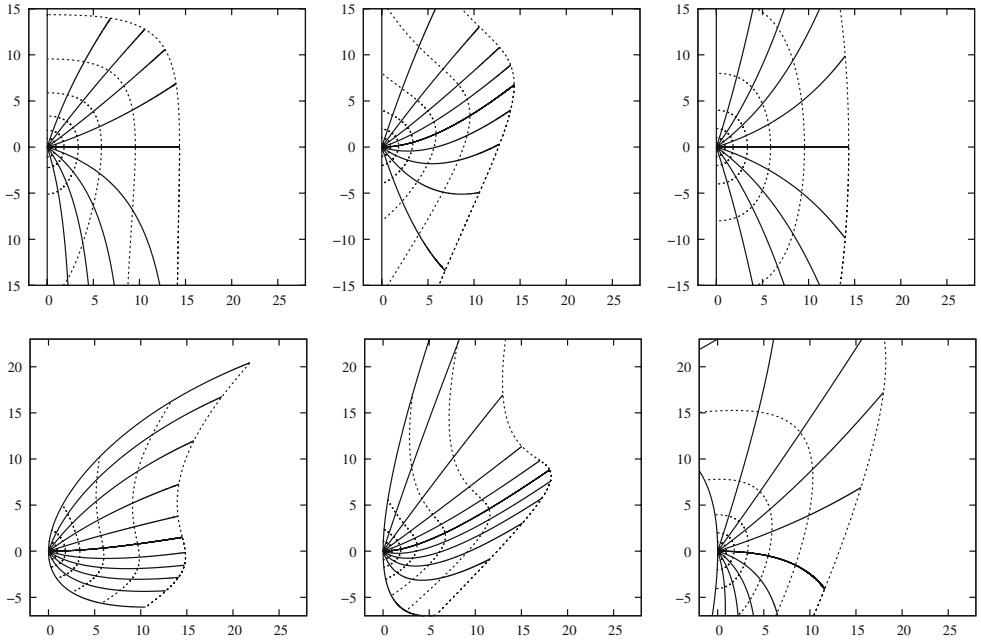

Fig. 6 Examples of the mappings predicted by a relaxation of the similarity hypothesis. The *top row* contains the new mappings defined in Sect. 4.6 without deformation (P is the identity matrix): from *left* to *right*, mapping 2. ($\lambda = 0.1$ and $\mu = 0.1$), mapping 3. ($\lambda = 0.1$ and $\nu = 0.1$), and mapping 4. ($\lambda = 0.1$). The *bottom row* represents mapping 1. (complex logarithmic, $\lambda = 0.1$), mapping 3. ($\lambda = 0.1$ and $\nu = 0.1$) and mapping 4. ($\lambda = 0.1$) with the deformation matrix $P = \begin{bmatrix} 1 & 0.5 \\ 0.2 & 1 \end{bmatrix}$. The *dashed lines* represent iso-amplitudes and *full lines*, iso-directions, as in Fig. 1. The axes units are millimeters and the same $B_x = 1.4$ mm and $B_y = 1.8$ mm parameters are used for all maps

As above, we perform the change of coordinates $S \mapsto \widetilde{S}$ to make $J(0) = I$. As in Sect. 4.2, using (Theorem 1, p. 225 and Proposition 7, p. 200 of Bourbaki 1972), guarantees the existence of two commuting matrices M_1 and M_2 such that

$$\widetilde{J}(\widetilde{S}) = \exp(M_1\ X + M_2\ Y) \tag{39}$$

By distinguishing between the different kind of sub-vector spaces $\mathbb{R}(M_1, M_2)$, we obtain five solutions, where P is a 2×2 invertible matrix (allowing a twist in the mapping) and $W = P^{-1}(\widetilde{S})$ is seen as a complex number $U + iV$.

1. $z = P\ \frac{1}{\lambda}\big(\exp(\lambda W) - 1\big) \quad \lambda \in \mathbb{C}$
2. $z = P \begin{pmatrix} \frac{1}{\lambda}\big(\exp(\lambda U) - 1\big) \\ \frac{1}{\mu}(\exp(\mu V) - 1) \end{pmatrix} \quad \lambda, \mu \in \mathbb{R}$
3. $z = P \begin{pmatrix} \frac{1}{\lambda}(\exp(\lambda U) - 1) \\ \exp(\lambda U)\big(V + \frac{\nu}{\lambda}\big(\frac{1}{\lambda} - U\big)\big) - \frac{\nu}{\lambda^2} \end{pmatrix} \quad \lambda, \nu \in \mathbb{R}$
4. $z = P \begin{pmatrix} \frac{1}{\lambda}(\exp(\lambda U) - 1) \\ V \end{pmatrix} \quad \lambda \in \mathbb{R}$
5. $z = \widetilde{S}$

Some examples of these mappings, with and without deformations are depicted in Fig. 6.

4.7 Description of the gluing simulation

Collicular maps are modeled by 90×90 tables of leaky-integrator neurons including 15 neurons borders. The system has the following hierarchical structure (see also Fig. 7):

1. The retina Ret encodes the target's position by a 2D Gaussian with standard deviation $\sigma = 2.5$, centered around the target's position.
2. The input layers inp receives the retinal signal with 70 ms delay. When the global activity passes a given threshold, it is transmitted to the motor layers (via a gluing mechanism which implements either ours or the Van Gisbergen's scheme) and the SBG OPNs are inhibited via LLBs.
3. The motor layers Mot send the command to the SBG, while their activities are integrated. When the integrator reaches a given threshold, the layers are inhibited and the saccade stops.
4. The SBG is first inhibited by the OPNs. The activity in Inp is transmitted to the LLBs which inhibit in turn the OPNs. When the activity in Inp is strong enough, OPNs are turned off and the EBNs/IBNs begin to receive the motor command from the mot layers through a weighted sum. This command is then integrated by the couple

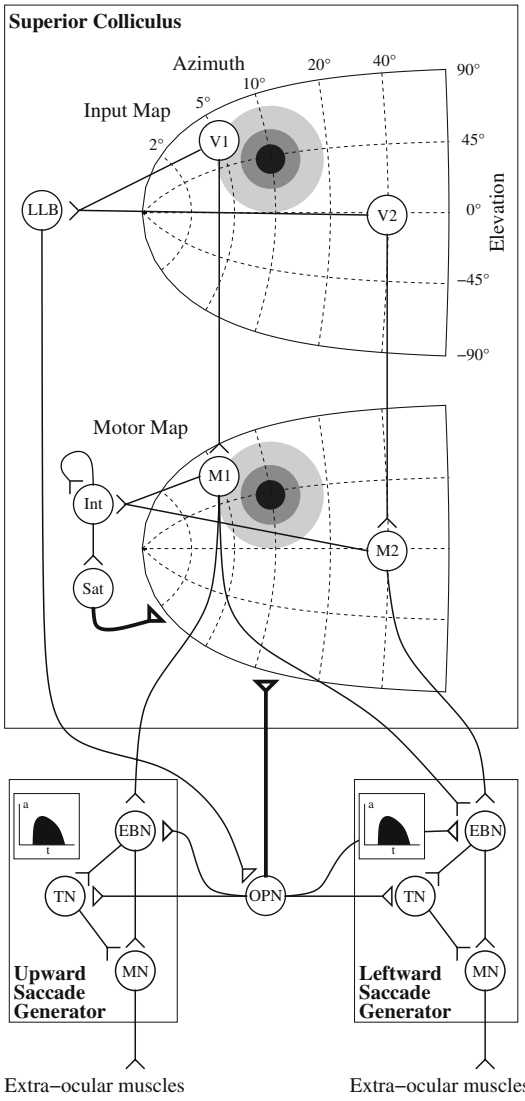

Fig. 7 Gluing simulation architecture. For simplicity, only one colliculus and two SBG (*upward and leftward*), without the crossed IBN projections, are represented. Moreover, only two neurons are represented in each collicular map (*V1, V2* and *M1, M2*, for visual and motor maps, respectively). *Shaded circles* in collicular maps represent the Gaussian activity generated by a (10°, 10°) target, while *insets* in the saccade generators represent the temporal code in the EBNs generated to drive the muscles. *Open triangles* represent excitatory synapses; *triangles* represent inhibitory synapses; *bold connections* affect the whole map. Refer to text for the abbreviations

of neurons TNs/MNs (tonic neurons/motoneurons). The activity of MNs is received by the eye plant (modeled by a second order differential equation) to generate the required eye's displacement.

The leaky-integrator rate neuron model used as building brick is as follows (τ: time constant in ms, I: input in mV):

$$\tau \frac{\mathrm{d}a}{\mathrm{d}t} = I - a \quad \text{and} \quad y = [a]^+$$

where the transfer function $[\ ]^+$ satisfies $[I]^+ = 0$ if $I < 0$ and $[I]^+ = I$ otherwise.

The input of Inp is

$$I_{\mathrm{Inp}^D}(S_0^D, t) = y_{\mathrm{Ret}^D}(S_0, t - t_0) \quad \text{with } D \in \{L, R\}$$

Long-Lead burst neurons (LLB), in charge of triggering saccades by inhibiting the OPN when the activity in the input layers reaches the ϵ_{trig} threshold, are modeled by the following:

$$I_{\mathrm{LLB}} = w_{\mathrm{Vis}}^{\mathrm{LLB}} \sum_S \left(y_{\mathrm{Inp}^R}(S) + y_{\mathrm{Inp}^L}(S)\right) - \epsilon_{\mathrm{trig}}$$

$$I_{\mathrm{OPN}} = -y_{\mathrm{LLB}} + \epsilon_{\mathrm{OPN}}$$

The activity in the motor layer Mot is gated by the OPNs and the integrating-saturating mechanism (note that saturation neurons have a longer time constant):

$$I_{\mathrm{Mot}^D}(S) = y_{\mathrm{Inp}^D}(S) - w_{\mathrm{OPN}}^{\mathrm{Mot}} y_{\mathrm{OPN}} - w_{\mathrm{Sat}}^{\mathrm{Mot}} y_{\mathrm{Sat}}$$

$$\text{with} \quad D \in \{R, L\}.$$

$$I_{\mathrm{Int}} = w_{\mathrm{Mot}}^{\mathrm{Int}} \sum_S \left(y_{\mathrm{Mot}^{\mathrm{R}}}(S) + y_{\mathrm{Mot}^{\mathrm{L}}}(S)\right)$$

$$I_{\mathrm{Sat}} = y_{\mathrm{Int}}(S) - \epsilon_{\mathrm{stop}}$$

The four SBG circuits (leftward, rightward, upward, downward) are identical, all of them are gated by OPN activity, and those operating in opposite directions are coordinated by the IBN crossed projections. The EBN and IBN activity is identical and defined by the following:

$$I_{\mathrm{BN}}^D = \sum_{X,Y} (w_\alpha y_{\mathrm{Mot}}(S)) - w_{\mathrm{OPN}}^{\mathrm{BN}} y_{\mathrm{OPN}}, \quad \text{for } D \in \{\mathrm{L}, \mathrm{R}\}$$

$$I_{BN}^D = \sum_{X,Y} \left(w_\beta y_{\mathrm{Mot}}(S)\right) - w_{\mathrm{OPN}}^{\mathrm{BN}} y_{\mathrm{OPN}}, \quad \text{for } D \in \{\mathrm{U}, \mathrm{D}\}$$

with the w_α and w_β defined in Eq. (3) of the main manuscript.

The tonic neurons are the only neurons modeled as perfect rather than leaky-integrators:

$$I_{\mathrm{TN}}^D = w_{\mathrm{BN}}^{\mathrm{TN}} \left(y_{\mathrm{EBN}}^D - y_{\mathrm{IBN}}^{D^{\mathrm{op}}}\right), \quad \text{with } D \in \{\mathrm{U}, \mathrm{D}, \mathrm{L}, \mathrm{R}\}$$

$$I_{\mathrm{MN}}^D = w_{\mathrm{BN}}^{\mathrm{MN}} \left(y_{\mathrm{EBN}}^D - y_{\mathrm{IBN}}^{D^{\mathrm{op}}}\right) + y_{\mathrm{TN}}, \quad \text{with } D \in \{\mathrm{U}, \mathrm{D}, \mathrm{L}, \mathrm{R}\}$$

where D^{op} is the opposite direction of D.

Table 1 Parameters of the model

τ	5 ms	τ_{Sat}	100 ms	t_0	70 ms
ϵ_{OPN}	100	ϵ_{trig}	400	ϵ_{stop}	200
$w_{\mathrm{Vis}}^{\mathrm{LLB}}$	0.005	$w_{\mathrm{OPN}}^{\mathrm{Mot}}$	40	$w_{\mathrm{OPN}}^{\mathrm{BN}}$	40
$w_{\mathrm{Mot}}^{\mathrm{Int}}$	0.002	$w_{\mathrm{Sat}}^{\mathrm{Mot}}$	8	$w_{\mathrm{BN}}^{\mathrm{TN}}$	0.05
$w_{\mathrm{BN}}^{\mathrm{MN}}$	1.52	w_{MN}^{θ}	4.07		

The eye plant model used is modeled as a second-order differential equation:

$$\ddot{\theta} + 0.6\dot{\theta} + 4\theta = w^{\theta}_{\mathrm{MN}} y_{\mathrm{MN}}$$

The parameters are summed up in the Table 1.

Acknowledgments The authors would like to gratefully thank A. Grantyn and A. Moschovakis for the valuable discussions, and H. Hicheur for valuable suggestions concerning the manuscript.

References

Anderson R, Keller E, Gandhi N, Das S (1998) Two-dimensional saccade-related population activity in superior colliculus in monkey. J Neurophysiol 80(2):798–817

Arai K, Keller E, Edelman J (1994) Two-dimensional neural network model of the primate saccadic system. Neural Netw 7:1115–1135

Arai K, Das S, Keller E, Aiyoshi E (1999) A distributed model of the saccade system: simulations of temporally perturbed saccades using position and velocity feedback. Neural Netw 12(10):1359–1375

Badler J, Keller E (2002) Decoding of a motor command vector from distributed activity in superior colliculus. Biol Cybern 86(3): 179–189

Bourbaki N (1972) Groupes et Algèbres de Lie, Chapitres 2 et 3. Dunod, Paris

Dräger U, Hugel D (1976) Topography of visual and somatosensory projections to mouse superior colliculus. J Neurophysiol 39:91–101

Feldon S, Feldon P, Kruger L (1970) Topography of the retinal projection upon the superior colliculus of the cat. Vision Res 10:135–143

Girard B, Berthoz A (2005) From brainstem to cortex: computational models of saccade generation circuitry. Prog Neurobiol 77:215–251

van Gisbergen J, van Opstal A, Tax A (1987) Collicular ensemble coding of saccades based on vector summation. Neuroscience 21(2):541–555

Goffart L, Pélisson D (1998) Orienting gaze shifts during muscimol inactivation of caudal fastigial nucleus in the cat. I. Gaze dysmetria. J Neurophysiol 79:1942–1958

Goossens H, van Opstal A (2000) Blink-perturbed saccades in monkey. II. superior colliculus activity. J Neurophysiol 83:3430–3452

Goossens H, van Opstal A (2006) Dynamic ensemble coding of saccades in the monkey superior colliculus. J Neurophysiol 95:2326–2341

Grantyn A, Moschovakis A (2003) Structure–function relationships in the superior colliculus of higher mammals. In: Hall W, Moschovakis V (eds) The superior colliculus: new approaches for studying sensorimotor integration, methods & new frontiers in neuroscience, chap 5. CRC Press, Boca Raton, pp 107–145

Grantyn A, Brandi AM, Dubayle D, Graf W, Ugolini G, Hadjidimitrakis K, Moschovakis A (2002) Density gradients of trans-synaptically labeled collicular neurons after injections of rabbies virus in the lateral rectus muscle of the rhesus monkey. J Comp Neurol 451:346–361

Groh J (2001) Converting neural signals from place codes to rate codes. Biol Cybern 85(3):159–165

Herrero L, Rodríguez F, Salas C, Torres B (1998) Tail and eye movemnts evoked by electrical microstimulation of the optic tectum in goldfish. Exp Brain Res 120:291–05

Hirsch H (1976) Differential topology. Springer, New York

Hörmander L (1983) The Analysis of linear partial differential operators I. No 256. In: Grundlehren der mathematischen Wissenschaften. Springer, Berlin

Iwamoto Y, Yoshida K (2002) Saccadic dysmetria following inactivation of the primate fastigial oculomotor region. Neurosci Lett 325:211–215

Kaneko C, Evinger C, Fuchs A (1981) Role of the cat pontine burst neurons in generation of saccadic eye movements. J Neurophysiol 46(3):387–408

Keller E (1974) Participation of medial pontine reticular formation in eye movement generation in monkey. J Neurophysiol 37(2): 316–332

King W, Fuchs F (1979) Reticular control of vertical saccadic eye movements by mesencephalic burst neurons. J Neurophysiol 42(3): 861–876

Knudsen E (1982) Auditory and visual maps of space in the optic tectum of the owl. J Neurosci 2(9):1177–1194

Lee C, Rohrer W, Sparks D (1988) Population coding of saccadic eye movements by neurons in the superior colliculus. Nature 332: 357–360

McIlwain J (1976) Large receptive fields and spatial transformations in the visual system. In: Porter R (eds) Neurophysiology II, Int Rev Physiol, vol 10. University Park Press, Baltimore, pp 223–248

McIlwain J (1983) Representation of the visual streak in visuotopic maps of the cat's superior colliculus: influence of the mapping variable. Vision Res 23(5):507–516

Moschovakis A, Kitama T, Dalezios Y, Petit J, Brandi A, Grantyn A (1998) An anatomical substrate for the spatiotemporal transformation. J Neurosci 18(23):10219–10229

Munoz D, Waitzman D, Wurtz R (1996) Activity of neurons in monkey superior colliculus during interrupted saccades. J Neurphysiol 75(6):2562–2580

Olivier E, Porter J, May P (1998) Comparison of the distribution and somatodendritic morphology of tectotecal neurons in the cat and monkey. Vis Neurosci 15:903–922

van Opstal A, van Gisbergen J (1989) A nonlinear model for collicular spatial interactions underlying the metrical properties of electrically elicited saccades. Biol Cybern 60(3):171–183

Optican L (2005) Sensorimotor transformation for visually guided saccades. Ann NY Acad Sci 1039:132–148

Ottes F, van Gisbergen JA, Eggermont J (1986) Visuomotor fields of the superior colliculus: a quantitative model. Vision Res 26(6): 857–873

Robinson D (1972) Eye movements evoked by collicular stimulation in the alert monkey. Vision Res 12:1795–1808

Rosa M, Schmid L (1994) Topography and extent of visual-field representation in the superior colliculus of the megachiropteran *Pteropus*. Vis Neurosci 11:1037–1057

Schwarz E (1980) Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. Vision Res 20:645–669

Siminoff R, Schwassmann H, Kruger L (1966) An electrophysiological study of the visual projection to the superior colliculus of the rat. J Comp Neurol 127:435–444

Soetedjo R, Kaneko C, Fuchs A (2000) Evidence that the superior colliculus participates in the feedback control of saccadic eye movements. J Neurophysiol 87:679–695

Sparks D, Holland R, Guthrie B (1976) Size and distribution of movement fields in the monkey superior colliculus. Brain Res 113:21–34

Yoshida K, McCrea R, Berthoz A, Vidal P (1982) Morphological and physiological characteristics of inhibitory burst neurons controlling horizontal rapid eye movements in the alert cat. J Neurophysiol 48(3):761–784

Système motivationnel adaptatif

A

A.1 Initialisation des ρ_i

Les paramètres initiaux ρ_{0i} de la fonction g ont été ajustés à la main afin de retrouver des performances comparables à celles du CBG original dans la tâche de survie (voir Tab. A.1).

A.2 Mise à jour des ρ_i

L'adaptation des ρ_i au cours d'une expérience s'effectue de la manière suivante : sur chaque période de 25 secondes consécutives, pour chaque type de source E ou E_p, un décompte du nombre de fois où une source de ce type entre dans le champ de vision de l'agent est effectué. On calcule en permanence des moyennes glissantes de ces décomptes sur les 100 dernières periodes, ce qui donne deux mesures de *disponibilité* des sources, a_E et a_{E_p}.

Ces mesures définissent des valeurs cibles de ρ_i, ρ_{Ti}, obtenues par des variations affines autour de ρ_{0i} :

$$\rho_{Ti} = \rho_{0i} + \beta_i^E(a_E - a_{0E}) + \beta_i^{Ep}(a_{E_p} - a_{0E_p}) \tag{A.1}$$

où a_{0E} et a_{0E_p} sont les disponibilités mesurées dans l'environnement de référence (1 E, 1E_p) sans adaptation (on trouve $a_{0E} = a_{0E_p} = 0.59$). Ainsi, dans cet environnement, on a $\rho_{Ti} = \rho_{0i}$, et dans des environnements plus ou moins riches en sources, le coefficient ρ_{Ti} est modifié en fonction des coefficients de proportionnalité β :

Action	β_i^E	$\beta_i^{E_p}$
ReloadE	0	0
ReloadEp	0	0
WanderE	0.25	0
WanderEp	0	0.25
Sleep	0.125	0.125
AvoidObstacle	0	0
ApproachE	0.25	0
ApproachEp	0	0.25

TAB. A.1 – ρ_0 *initiaux.*

Action	ρ_0
ReloadE	0.98
ReloadEp	0.95
WanderE	0.88
WanderEP	0.88
Sleep	0.60
AvoidObstacle	0.95
ApproachE	0.71
ApproachEp	0.71

Les valeurs ρ_i sont alors adaptées à chaque pas de temps vers les valeurs ρ_{Ti} :

$$\rho_i \leftarrow \rho_i + \alpha(\rho_{Ti} - \rho_i) \tag{A.2}$$

Dans nos expériences, le coefficient d'apprentissage vaut $\alpha = 0.002$ unités par seconde.

Curriculum Vitae

B

Benoît Girard
CR1 CNRS

3 rue Leneveux — benoit.girard@isir.fr
75014 Paris — 01 44 27 63 81
Né le 9 nov. 1975 (Paris 14) — Nationalité Française
Entrée au CNRS : 01/09/2005 — Agent : 00035560

Formation

2000-2003 **Thèse en Informatique.** Université Pierre et Marie Curie (UPMC).
Intégration de la navigation et de la sélection de l'action dans une architecture de contrôle inspirée des ganglions de la base.
Sous la direction de A. Guillot et A. Berthoz.

1998-2000 **DEA IARFA** (Intelligence Artificielle, Reconnaissance des Formes et Applications) de l'UPMC, section Vie Artificielle, mention Bien.

1995-1998 **Ingénieur ECN** (Ecole Centrale Nantes), option Informatique, mention Bien.

Expérience

2009-présent **Chargé de recherche (CR1 - CID 44).**
Institut des systèmes intelligents et de robotique (ISIR - UMR 7222, CNRS - UPMC),
équipe *Systèmes Intégrés Mobiles et Autonomes (SIMA).*

2005-2008 **Chargé de recherche (CR2 - CID 44).**
Laboratoire de Physiologie de la Perception et de l'Action (LPPA - UMR 7152, CNRS - Collège de France),
équipe *Mémoire spatiale et contrôle du mouvement.*

2003-2005 **Post-Doctorat** au LPPA (UMR 7124, CNRS - Collège de France).
Modèle computationnel contractant du système saccadique : du tronc cérébral au cortex.

Enseignements & Encadrement

C

Activités d'enseignement (depuis 2005)

2009-2010

- **University of British Columbia (UBC, Vancouver, Canada)**
  - *Neural substrate of ocular movements*, cours, Master , 1h30.

2008-2010

- **UPMC**
  - *Traitement de l'information pour la sélection de l'action*, cours, M2 Biologie Intégrative et Physiologie (BIP), 3h.
  - *Modélisation des stratégies de navigation et de leurs interactions*, cours, M2 BIP, 3h.

2006-2010

- **UPMC**
  - *Sélection de l'action*, cours, M2 Intelligence Artificielle et Décision (IAD), 3h.

2006-2009

- **UPMC**
  - *Bases neurales du contrôle du regard*, cours, M2 BIP, 2h.

2007-2008

- **Ecole des Hautes Etudes en Sciences Sociales (EHESS).**
  - *Mouvements des yeux : Neurophysiologie*, cours, M2 Sciences Cognitives (CogMaster), 3h

2005-2006

- **UPMC**
  - *Modélisation neuromimétique de la sélection de l'action : les ganglions de la base*, cours, M2 IAD, 2h.

Activités d'encadrement

2009–...

- **Jean Liénard**, Université Pierre et Marie Curie (UPMC), doctorat, co-direction A. Guillot (HdR)
  Evolution artificielle de modèles neuromimétiques de sélection de l'action
- **David Tlalolini-Romero**, post-doctorat, durée 14 mois, co-encadrement avec A. Berthoz
  Locomotion humanoïde bio-inspirée
- **Mariella Dimiccoli**, post-doctorat, durée 12 mois, co-encadrement avec A. Berthoz et D. Bennequin
  Modélisation du rôle de la géométrie dans la fonction du système vestibulaire
- **Charles Thurat**, UPMC, stage de M2 de Biologie Intégrative et Physiologie, durée 6 mois,
  Modélisation des mécanismes de sélection dans les boucles tecto-basales.

2008–2009

- **Jean Liénard**, Université Paris Sud (Paris XI) & Ecole Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise (ENSIIE), stage de M2 d'Informatique et de 3ème année d'école d'ingénieur, durée 6 mois
  Mise au point d'un modèle neuromimétique de sélection de stratégie de navigation en environnement simulé
- **Cécile Masson**, Polytech Paris Sud, stage de M2 de Sciences Cognitives et de 3ème année d'école d'ingénieur, durée 5 mois
  Modélisation de l'intégration de chemin chez le rat à partir des cellules de grilles

2007–2008

- **Alexandre Coninx**, Ecole des Hautes Etudes en Sciences Sociales (EHESS), stage de M2 de Sciences Cognitives, durée 5 mois, co-direction A. Guillot
  Modulation motivationnelle adaptative dans un modèle des ganglions de la base pour la sélection de l'action
- **Charles Thurat**, Ecole Normale Supérieure de Cachan, stage de M1, durée 2 mois
  Etude de la paramétrisation d'un modèle des ganglions de la base
- **Francis Colas**, post-doctorat, durée 18 mois
  Modélisation bayesienne des processus de sélection de cible dans une tâche de MOT

2006

- **Fabien Flacher**, post-doctorat, durée 6 mois
  Modélisation bayesienne des processus de sélection de cible dans une tâche de MOT

2004 ; 2005–2008

- **Nicolas Tabareau**, Ecole Normale Supérieure de Cachan, stage de M1, durée 2 mois
  Utilisation de la contraction en neurosciences computationnelles (check)
  Après la fin officielle de ce stage, j'ai continué à encadrer les travaux de N. Tabareau au LPPA (à raison d'un jour par semaine en moyenne), alors qu'il poursuivait en parallèle son M2 puis sa thèse en Informatique au laboratoire Preuve, Programmes et Systèmes (PPS, UMR7126) . Il a, pendant cette période (2005–2008), produit un travail du niveau d'une thèse en sciences cognitives, comme en témoignent ses publications (Girard et al., 2005b, 2006a; Manfredi et al., 2006; Tabareau et al., 2007; Girard et al., 2008).

2003

- **Mehdi Khamassi**, Université Pierre et Marie Curie (UPMC), stage de DEA de Sciences Cognitives, durée 6 mois, co-encadrement A. Guillot
  Un modèle d'apprentissage par renforcement dans une architecture de contrôle de la sélection de l'action chez le rat artificiel Psikharpax

2002

- **Sébastien Laithier**, Université Pierre et Marie Curie (UPMC), stage de DEA Intelligence Artificielle, Reconnaissance des Formes et Applications, durée 6 mois, co-encadrement A. Guillot
  Comparaison de mécanismes de sélection de l'action pour un robot Lego

Publications

D

Listes des publications principales, approuvées par des comités de lecture, tout d'abord dans les journaux scientifiques (10), puis dans les conférences (16).

D.1 Journaux à comité de lecture

- L. Dollé, D. Sheynikhovich, **B. Girard**, R. Chavarriaga, A. Guillot (2010). Path planning versus cue responding : a bioinspired model of switching between navigation strategies. *Biological Cybernetics*, 103(4) :299-317.
- F. Colas, F. Flacher, T. Tanner, P. Bessière and **B. Girard** (2009). Bayesian models of eye movement selection with retinotopic maps. *Biological Cybernetics*, 100(3) :203-214.
- **B. Girard**, N. Tabareau, Q.C. Pham, A. Berthoz and J.-J. Slotine (2008). Where neuroscience and dynamic system theory meet autonomous robotics : a contracting basal ganglia model for action selection. *Neural Networks*, 21(4) :628-641.
- N. Tabareau, D. Bennequin, A. Berthoz, J.-J. Slotine and **B. Girard** (2007). Geometry of the superior colliculus mapping and efficient oculomotor computation. *Biological Cybernetics*, 97(4) :279-292.
- **B. Girard** and A. Berthoz (2005). From brainstem to cortex : computational models of the saccade generation circuitry. *Progress in Neurobiology*. 77(4) :215-251.
- **B. Girard**, D. Filliat, J.-A. Meyer, A. Berthoz and A. Guillot (2005). Integration of navigation and action selection in a computational model of cortico-basal ganglia-thalamo-cortical loops. *Adaptive Behavior*. 13(2) :115-130.
- M. Khamassi, L. Lachèze, **B. Girard**, A. Berthoz and A. Guillot (2005). Actor-critic models of reinforcement learning in the basal ganglia : From natural to artificial rats. *Adaptive Behavior*, 13(2) : 131-148.
- J.-A. Meyer, A. Guillot, **B. Girard**, M. Khamassi, P. Pirim and A. Berthoz (2005). The Psikharpax project : Towards building an artificial rat. *Robotics and Autonomous Systems*, 50(4) :211-223.
- **B. Girard**, V. Cuzin, A. Guillot, K.N. Gurney and T.J. Prescott (2003). A Basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of Integrative Neuroscience*, 2(2) :179-200.
- **B. Girard**, G. Robert and A. Guillot (2001). Jeux Vidéo et Intelligence Artificielle Située. *In Cognito* 22 :57-72.

D.2 Actes de conférence à comité de lecture

- J.-B., Mouret, S., Doncieux, **B. Girard** (2010). Importing the Computational Neuroscience Toolbox into Neuro-Evolution—Application to Basal Ganglia. In *GECCO'10 : Proceedings of the 12th annual conference on Genetic and Evolutionary Computation*. ACM.
- J. Liénard, A. Guillot, **B. Girard** (2010). Multi-Objective Evolutionary Algorithms to Investigate Neurocomputational Issues : The Case Study of Basal Ganglia Models. In *From Animals to Animats 11 : Proceedings of the Eleventh International Conference on Simulation of Adaptive Behavior*, Lecture Notes in Artificial Intelligence, pages 602–609. Springer.
- S. N'Guyen, P. Pirim, J.-A. Meyer, **B. Girard** (2010). An Integrated Neuromimetic Model of the Saccadic Eye Movements for the Psikharpax Robot. In *From Animals to Animats 11 : Proceedings of the Eleventh International Conference on Simulation of Adaptive Behavior*, Lecture Notes in Artificial Intelligence, pages 115–126. Springer.
- L. Dollé, D. Sheynikhovich, **B. Girard**, B. Ujfalussy, R. Chavariagga, A. Guillot (2010). Analyzing interactions between cue-guided and place-based navigation with a computational model of action selection : Influence of sensory cues and training. In *From Animals to Animats 11 : Proceedings of the Eleventh International Conference on Simulation of Adaptive Behavior*, Lecture Notes in Artificial Intelligence, pages 338–349. Springer.
- C. Masson, **B. Girard** (2009). Decoding the grid cells for metric navigation using the residue numeral system. In *2nd International Conference on Cognitive Neurodynamics (ICCN2009)*, Hangzhou, China.
- M.-T. Tran, P. Souères, M. Taïx, **B. Girard** (2009). Eye-centered vs body-centered reaching control : A robotics insight into the neuroscience debate. *Robotics and Biomimetics (ROBIO 2009)*.
- L. Dollé, M. Khamassi, **B. Girard**, A. Guillot and R. Chavariagga (2008). Analyzing interactions between navigation strategies using a computational model of action selection. In *Spatial Cognition VI*, Lecture Notes in Computer Science, pages 71-86. Springer.
- A. Coninx, A. Guillot and **B. Girard** (2008). Adaptive motivation in a biomimetic action selection mechanism. In *NeuroComp 2008*, pages 158-162.
- F. Colas, F. Flacher, P. Bessière and **B. Girard** (2008). Explicit uncertainty for eye movement selection. In *NeuroComp 2008*, pages 103-107.
- **B. Girard**, N. Tabareau, A. Berthoz and J.-J. Slotine (2006). Selective amplification using a contracting model of the basal ganglia. In Alexandre, F., Boniface, Y., Bougrain, L., Girau, B. and Rougier, N. (Eds), *NeuroComp 2006*, pages 30-33.
- L. Manfredi, E. Maini, C. Laschi, P. Dario, **B. Girard**, N. Tabareau and A. Berthoz, A. (2006). Implementation of a neurophysiologic model of saccadic movements on an anthropomorphic robotic head. In *IEEE-RAS Int. Conf. on Humanoid Robots*, pages 438-443.
- **B. Girard**, N. Tabareau, J.J. Slotine and A. Berthoz (2005). Contracting model of the basal ganglia. In Bryson, J., Prescott, T. and Seth,

A. (Eds) *Modelling Natural Action Selection : Proceedings of an International Workshop*, pages 69-76. AISB Press, Brighton, UK.

- M. Khamassi, **B. Girard**, A. Berthoz and A. Guillot (2004). Comparing three critic models of reinforcement learning in the basal ganglia connected to a detailed actor in a S-R task. In Groen, F., Amato, N., Bonarini, A., Yoshida, E., and Kröse, B. (Eds), *Proceedings of the Eighth International Conference on Intelligent Autonomous Systems*, pages 430-437. IOS Press, Amsterdam, The Netherlands.
- D. Filliat, **B. Girard**, A. Guillot, M. Khamassi, L. Lachèze and J.-A. Meyer (2004). State of the artificial rat Psikharpax. In Schaal, S., Ijspeert, A., Billard, A. , Vijayakumar, S., Hallam, J., and Meyer, J.-A. (Eds), *From Animals to Animats 8 : Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, pages 3-12. MIT Press, Cambridge, MA.
- **B. Girard**, D. Filliat, J.-A. Meyer, A. Berthoz and A. Guillot (2004). An integration of two control architectures of action selection and navigation inspired by neural circuits in the vertebrates : The basal ganglia. In Bowman, H. and Labiouse, C. (Eds), *Connectionist Models of Cognition and Perception II, Proceedings of the Eighth Neural Computation and Psychology Workshop*, pages 72-81. World Scientific, Singapore.
- **B. Girard**, V. Cuzin, A. Guillot, K.N. Gurney and T.J. Prescott (2002). Comparing a bio-inspired robot action selection mechanism with winner-takes-all. *From Animals to Animats 7. Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, pages 75-84. The MIT Press.

Projets de recherche

E

Les travaux présentés dans ce document ont pour la plupart été réalisés dans le cadre de projets financés par l'Union Européenne, le CNRS ou l'ANR. La liste suivante répertorie ces projets et les travaux qui leurs sont rattachés :

- **Psikharpax** (Filliat et al., 2004; Meyer et al., 2005), financé par le programme ROBEA du CNRS de 2001 à 2004, qui visait à élaborer un "rat artificiel", c'est-à-dire synthétiser dans un robot les mécanismes adaptatifs et les structures nerveuses connus pour être impliqués dans la navigation et la sélection de l'action chez le rat.

  Travail de recherche : section(s) 2.2.1
- **ICEA** (Integrating Cognition Emotion and Autonomy), financement européen IST-027819 de 2006 à 2009, qui fait suite à Psikharpax, et qui vise à concevoir des architectures de contrôle inspirées du cerveau des mammifères, intégrant des processus cognitifs, émotionnels et de régulation interne, incarnées dans des plate-formes robotiques.

  Travail de recherche : section(s) 2.2.2, 3.1 & 3.2
- **NEUROBOTICS** (The fusion of Neuroscience and Robotics), financement européen FP6-IST-001917 de 2004 à 2007, qui visait à fusionner neurosciences et robotique pour investiguer le domaine des systèmes bioniques hybrides.

  Travail de recherche : section(s) 2.1 & 4.1
- **BIBA** (Bayesian Inspired Brain and Artefacts), financement européen IST-2001-32115 de 2002 à 2005, qui visait à utiliser la logique probabiliste pour comprendre le fonctionnement du cerveau et implémenter des agents intelligents.

  Travail de recherche : section(s) 4.1
- **BACS** (Bayesian Approach to Cognitive Systems), financement européen FP6-IST-027140 de 2006 à 2010, qui fait suite à BIBA et poursuit des objectifs similaires.

  Travail de recherche : section(s) 2.3
- **ROMA** (Représentation Oculocentrée et Mouvements d'Atteinte), financé par le programme Neuro-Informatique du CNRS de 2007 à 2009, qui vise à explorer les liens entre la commande référencée capteur en robotique et la génération de mouvements d'atteinte chez le primate.

  Travail de recherche : section(s) 5.1.1
- **ROMEO** (Robot humanoïde compagnon et assistant personnel), projet industriel avec la société Aldébaran visant à la conception

d'un robot humanoïde de grande taille (>1m20) pour usage domestique.

- **CLONS** (CLOsed-loop Neural prostheses for vestibular disorderS), financement européen FP7-ICT-225929 de 2009 à 2012, qui vise à développer une prothèse de système vestibulaire implantable.
- **EvoNeuro** (Evolution Artificielle et Neurosciences Computationnelles), financement ANR ANR-09-EMER-005-01 de 2009 à 2012, qui vise à explorer les fertilisations croisées de l'évolution artificielle et des neurosciences computationnelles.

Travail de recherche : section(s) 5.2

Bibliographie

S. Albertin, A. B. Mulder, E. Tabuchi, M. B. Zugaro, et S. I. Wiener. Lesion of the medial shell of the nucleus accumbens impair rats in finding larger rewards, but spare reward-seeking behavior. *Behavioural Brain Research*, 117 :173–183, 2000. (Cité page 24.)

G. E. Alexander, M. D. Crutcher, et M. R. DeLong. Basal ganglia-thalamocortical circuits : Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Progress in Brain Research*, 85 :119–146, 1990. (Cité page 7.)

G. E. Alexander, M. R. DeLong, et P. L. Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9 :357–381, 1986. (Cité page 7.)

K. Arai, S. Das, E.L. Keller, et E. Aiyoshi. A distributed model of the saccade system : simulations of temporally perturbed saccades using position and velocity feedback. *Neural Netw*, 12(10) :1359–1375, Dec 1999. (Cité page 59.)

M. A. Arbib et P. F. Dominey. Modeling the roles of basal ganglia in timing and sequencing saccadic eye movements. Dans J. C. Houk, J. L. Davis, et D. G. Beiser, éditeurs, *Models of Information Processing in the Basal Ganglia*, pages 149–162. The MIT Press, Cambridge, MA, 1995. (Cité page 58.)

A. Arleo et L. Rondi-Reig. Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms. *J Integr Neurosci*, 6(3) :327–366, Sep 2007. (Cité page 37.)

G. Baldassarre. A modular neural-network model of the basal ganglia's role in learning and selecting motor behavior. *Journal of Cognitive Systems Research*, 3(1) :5–13, 2002. (Cité page 24.)

C. Barry, R. Hayman, N. Burgess, et K.J. Jeffery. Experience-dependent rescaling of entorhinal grids. *Nat Neurosci*, 10(6) :682–684, Jun 2007. URL `http://dx.doi.org/10.1038/nn1905`. (Cité page 47.)

A. G. Barto. Adaptive critics and the basal ganglia. Dans J. C. Houk, J. L. Davis, et D. G. Beiser, éditeurs, *Models of Information Processing in the Basal Ganglia*, pages 215–232. The MIT Press, Cambridge, MA, 1995. (Cité page 8.)

M.A. Basso et R.H. Wurtz. Modulation of neuronal activity in superior colliculus by changes in target probability. *J Neurosci*, 18(18) :7519–34, 1998. (Cité page 11.)

D. G. Beiser et J. C. Houk. Model of cortical-basal ganglionic processing : encoding the serial order of sensory events. *Journal of Neurophysiology*, 79 :3168–3188, 1998. (Cité page 17.)

K. C. Berridge et T. E. Robinson. What is the role of dopamine in reward : hedonic impact, reward learning , or incentive salience. *Brain Research Reviews*, 28 :309–369, 1998. (Cité pages 8 et 31.)

P. Bessière, J.-M. Ahuactzin, O. Aycard, D. Bellot, F. Colas, Ch. Coué, J. Diard, R. Garcia, C. Koike, O. Lebeltel, R. Le Hy, O. Malrait, E. Mazer, K. Meknacha, C. Pradalier, et A. Spalanzani. Survey : Probabilistic methodology and techniques for artefact conception and development. Rapport Technique 4730, INRIA, 2003. (Cité page 31.)

P. Bessière, Ch. Laugier, et R. Siegwart. *Probabilistic reasoning and decision making in sensory-motor systems*, volume 46 de *Tracts in Advanced Robotics*. Springer, 2008. (Cité page 31.)

M.D. Bevan, P.A. Booth, S.A. Eaton, et J.P. Bolam. Selective innervation of neostriatal interneurons by a subclass of neurons in the globus pallidus of rats. *Journal of Neuroscience*, 18(22) :9438–9452, 1998. (Cité page 17.)

R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1) :14–23, 1986. (Cité page 1.)

J.J. Bryson. Mechanisms of action selection : Introduction to the special issue. *Adaptive Behavior*, 15(1) :5–8, 2007. URL `http://adb.sagepub.com/cgi/reprint/15/1/5`. (Cité page 13.)

J.J.. Bryson, T.J. Prescott, et A.K. Seth, éditeurs. *Modeling Natural Action Selection*, Edinburgh, UK, 2005. (Cité page 13.)

N. Burgess. Spatial cognition and the brain. *Ann N Y Acad Sci*, 1124 :77–97, Mar 2008. URL `http://dx.doi.org/10.1196/annals.1440.002`. (Cité page 37.)

C.D. Carello et R.J. Krauzlis. Manipulating intent : evidence for a causal role of the superior colliculus in target selection. *Neuron*, 43(4) :575–583, Aug 2004. URL `http://dx.doi.org/10.1016/j.neuron.2004.07.026`. (Cité page 11.)

R. Chavarriaga, T. Strösslin, D. Sheynikhovich, et W. Gerstner. A Computational Model of Parallel Navigation Systems in Rodents. *Neuroinformatics*, 3(3) :223–242, 2005. (Cité pages 39, 40, 44 et 48.)

G. Chevalier et M. Deniau. Disinhibition as a basic process of striatal functions. *Trends in Neurosciences*, 13 :277–280, 1990. (Cité pages 6 et 7.)

H. J. Chiel et R. D. Beer. The brain has a body : adaptive behavior emerges from interactions of nervous system, body and environment. *Trends Neurosci*, 20(12) :553–557, Dec 1997. (Cité page 1.)

F. Colas, J. Droulez, M. Wexler, et P. BessiÃ¨re. A unified probabilistic model of the perception of three-dimensional structure from optic flow. *Biol Cybern*, 97(5-6) :461–477, Dec 2007. URL `http://dx.doi.org/10.1007/s00422-007-0183-z`. (Cité page 31.)

F. Colas, F. Flacher, P. Bessière, et B. Girard. Explicit uncertainty for eye movement selection. Dans E. Daucé et L. Perrinet, éditeurs, *NeuroComp 2008*, pages 103–107, Marseille, France, 2008. URL http://isir.robot.jussieu.fr/webadmin/files/2008ACTN856.pdf. ISBN : 978-2-9532965-0-1. (Cité pages 3 et 31.)

F. Colas, F. Flacher, T. Tanner, P. Bessière, et B. Girard. Bayesian models of eye movement selection with retinotopic maps. *Biol Cybern*, 100(3) :203–214, Mar 2009. URL http://dx.doi.org/10.1007/s00422-009-0292-y. (Cité pages 3 et 31.)

A. Coninx, A. Guillot, et B. Girard. Adaptive motivation in a biomimetic action selection mechanism. Dans Laurent U. Perrinet et Emmanuel Daucé, éditeurs, *Proceedings of the second french conference on Computational Neuroscience*, pages 158–162, Marseille, France, 2008. URL http://2008.neurocomp.fr. ISBN : 978-2-9532965-0-1. (Cité pages 3, 23, 27 et 30.)

N.D. Daw, Y. Niv, et P. Dayan. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci*, 8(12) :1704–1711, 2005. (Cité page 48.)

S. Denève. Bayesian spiking neurons i : inference. *Neural Comput*, 20 (1) :91–117, Jan 2008a. URL http://dx.doi.org/10.1162/neco.2008.20.1.91. (Cité page 35.)

S. Denève. Bayesian spiking neurons ii : learning. *Neural Comput*, 20 (1) :118–145, Jan 2008b. URL http://dx.doi.org/10.1162/neco.2008.20.1.118. (Cité page 35.)

B. D. Devan et N. M. White. Parallel information processing in the dorsal striatum : relation to hippocampal function. *J Neurosci*, 19(7) :2789–2798, Apr 1999. (Cité pages 37, 40 et 42.)

L. Dollé, M. Khamassi, B. Girard, A. Guillot, et R. Chavarriaga. Analyzing interactions between navigation strategies using a computational model of action selection. Dans *Proceedings of the international conference on Spatial Cognition VI : Learning, Reasoning, and Talking about Space*, pages 71–86. Springer, 2008. (Cité pages 4 et 38.)

L. Dollé, D. Sheynikhovich, B. Girard, R. Chavarriaga, et A. Guillot. Path planning versus cue responding : a bioinspired model of switching between navigation strategies. *Biological Cybernetics*, 103(4) :299–317, 2010a. (Cité pages 4, 38, 39, 40 et 48.)

L. Dollé, D. Sheynikhovich, B. Girard, B. Ujfalussy, R. Chavariagga, et A. Guillot. Analyzing interactions between cue-guided and place-based navigation with a computational model of action selection : Influence of sensory cues and training. Dans *From Animals to Animats 11 : Proceedings of the Eleventh International Conference on Simulation of Adaptive Behavior*, LNAI, pages 338–349. Springer, 2010b. (Cité pages 4, 38 et 39.)

P. Dominey, M. Arbib, et J.-P. Joseph. A model of corticostriatal plasticity for learning oculomotor associations and sequences. *Journal of Cognitive Neuroscience*, 7 :311–336, 1995. (Cité page 58.)

P. F. Dominey et M. A. Arbib. A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex*, 2 :153–175, 1992. (Cité page 58.)

K. Doya. Modulators of decision making. *Nat Neurosci*, 11(4) :410–416, Apr 2008. URL http://dx.doi.org/10.1038/nn2077. (Cité page 13.)

K. Doya, K. Samejima, K.-I. Katagiri, et M. Kawato. Multiple model-based reinforcement learning. *Neural Comput*, 14(6) :1347–1369, Jun 2002. URL http://dx.doi.org/10.1162/089976602753712972. (Cité page 24.)

J. Droulez et A. Berthoz. A neural network model of sensoritopic maps with predictive short-term memory properties. *Proceedings of the National Academy of Science*, 88 :9653–9657, 1991. (Cité page 32.)

P. Dupuis et A. Nagurney. Dynamical systems and variational inequalities. *Annals of Operations Research*, 44(1) :7–42, 1993. (Cité page 15.)

A. Elfes. *Occupancy grids : a probabilistic framework for robot perception and navigation*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 1989. (Cité page 32.)

A.S. Etienne et K.J. Jeffery. Path integration in mammals. *Hippocampus*, 14(2) :180–192, 2004. URL http://dx.doi.org/10.1002/hipo.10173. (Cité pages 45 et 48.)

G. Felsen et Z.F. Mainen. Neural substrates of sensory-guided locomotor decisions in the rat superior colliculus. *Neuron*, 60(1) :137–148, Oct 2008. URL http://dx.doi.org/10.1016/j.neuron.2008.09.019. (Cité pages 9 et 43.)

I.R. Fiete, Y. Burak, et T. Brookings. What grid cells convey about rat location. *J Neurosci*, 28(27) :6858–6871, Jul 2008. URL http://dx.doi.org/10.1523/JNEUROSCI.5684-07.2008. (Cité pages 46 et 47.)

D. Filliat, B. Girard, A. Guillot, M. Khamassi, L. Lachèze, et J.-A. Meyer. State of the artificial rat Psikharpax. Dans S. Schaal, A. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam, et J.-A. Meyer, éditeurs, *From Animals to Animats 8*, pages 2–12. MIT Press, Cambridge, MA, 2004. (Cité page 157.)

D. J. Foster, R. G. Morris, et P. Dayan. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus*, 10(1) :1–16, 2000. URL http://dx.doi.org/3.0.CO;2-1. (Cité page 47.)

B. Girard et A. Berthoz. From brainstem to cortex : Computational models of saccade generation circuitry. *Progress in Neurobiology*, 77 :215–251, 2005. (Cité page 58.)

B. Girard, V. Cuzin, A. Guillot, K. N. Gurney, et T. J. Prescott. Comparing a bio-inspired robot action selection mechanism with winner-takes-all. Dans B. Hallam, D. Floreano, J. Hallam, G. Hayes, et J.-A. Meyer, éditeurs, *From Animals to Animats 7 : Proceedings of the Seventh International*

Conference on Simulation of Adaptive Behavior, pages 75–84. The MIT Press, Cambridge, MA, 2002. (Cité pages 3 et 14.)

B. Girard, V. Cuzin, A. Guillot, K. N. Gurney, et T. J. Prescott. A basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of Integrative Neuroscience*, 2(2) :179–200, 2003. (Cité pages 3, 13, 14, 19, 20 et 22.)

B. Girard, D. Filliat, J.-A. Meyer, A. Berthoz, et A. Guillot. An integration of two control architectures of action selection and navigation inspired by neural circuits in the vertebrate : the basal ganglia. Dans H. Bowman et C. Labiouse, éditeurs, *Connectionist Models of Cognition and Perception II*, volume 15 de *Progress in Neural Processing*, pages 72–81. World Scientific, Singapore, 2004. (Cité pages 3, 4 et 37.)

B. Girard, D. Filliat, J.-A. Meyer, A. Berthoz, et A. Guillot. Integration of navigation and action selection in a computational model of cortico-basal ganglia-thalamo-cortical loops. *Adaptive Behavior*, 13(2) :115–130, 2005a. (Cité pages 3, 4, 13, 14, 22, 37, 39, 48 et 57.)

B. Girard, N. Tabareau, A. Berthoz, et J.-J. Slotine. Selective amplification using a contracting model of the basal ganglia. Dans F. Alexandre, Y. Boniface, L. Bougrain, B. Girau, et N. Rougier, éditeurs, *NeuroComp 2006*, pages 30–33, 2006a. (Cité pages 14 et 151.)

B. Girard, N. Tabareau, Q.C. Pham, A. Berthoz, et J.-J. Slotine. Where neuroscience and dynamic system theory meet autonomous robotics : a contracting basal ganglia model for action selection. *Neural Networks*, 21 (4) :628–641, 2008. (Cité pages 3, 14, 15, 17, 22, 58, 151 et 173.)

B. Girard, N. Tabareau, J.-J. Slotine, et A. Berthoz. Contracting model of the basal ganglia. Dans J. Bryson, T. Prescott, et A. Seth, éditeurs, *Modelling Natural Action Selection : Proceedings of an International Workshop*, pages 69–76, Brighton, UK, 2005b. AISB Press. (Cité pages 3, 14 et 151.)

B. Girard, N. Tabareau, J.-J. Slotine, et A. Berthoz. Using contraction analysis to design a model of the cortico-baso-thalamo-cortical loops. Dans A.K. Ijspeert, J. Buchli, A. Selverston, M. Rabinovitch, M. Hasler, W. Gerstner, A. Billard, H. Markram, et D. Floreano, éditeurs, *EPFL LATSIS Symposium 2006, Dynamical principles for neuroscience and intelligent biomimetic devices*, pages 85–86, Lausanne, Switzerland, 2006b. EPFL. (Cité page 3.)

H.H.L.M. Goossens et A.J. van Opstal. Blink-perturbed saccades in monkey. II. superior colliculus activity. *Journal of Neurophysiology*, 83 :3430–3452, 2000. (Cité page 55.)

H.H.L.M. Goossens et A.J. van Opstal. Dynamic ensemble coding of saccades in the monkey superior colliculus. *J. Neurophy.*, 95 :2326–2341, 2006. (Cité pages 52, 53 et 55.)

A.A. Grantyn et A. Moschovakis. Structure-function relationships in the superior colliculus of higher mammals. Dans W.C. Hall et A. Moschovakis, éditeurs, *The superior colliculus : new approaches for studying sensori-*

motor integration, Methods & new frontiers in neuroscience, Chapitre 5, pages 107–145. CRC Press, Boca Raton, FL., 2003. (Cité page 55.)

J.M. Groh. Converting neural signals from place codes to rate codes. *Biol. Cybern.*, 85(3) :159–165, 2001. (Cité pages 52 et 53.)

A. Guazzelli, F. J. Corbacho, M. Bota, et M. A. Arbib. Affordances, motivations and the worls graph theory. *Adaptive Behavior*, 6(3/4) :435–471, 1998. (Cité pages 39 et 48.)

A. Guillot et J.-A. Meyer. From SAB94 to SAB2000 : What's new, animat ? Dans J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, et S. W. Wilson, éditeurs, *From Animals to Animats 6 : Proceedings of the sixth international conference on simulation of adaptive behavior*, pages 3–12. MIT Press, 2000. (Cité page 1.)

A. Guillot et J.-A. Meyer. The animat contribution to cognitive systems research. *Journal of Cognitive Systems Research*, 2(2) :157–165, 2001. (Cité page 1.)

K. Gurney, T. J. Prescott, et P. Redgrave. A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, 84 :401–410, 2001a. (Cité pages 13, 14 et 173.)

K. Gurney, T. J. Prescott, et P. Redgrave. A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biological Cybernetics*, 84 :411–423, 2001b. (Cité pages 13, 14, 18, 19, 23 et 173.)

K.N. Gurney. Reverse engineering the vertebrate brain : Methodological principles for a biologically grounded programme of cognitive modelling. *Cognitive Computation*, 1 :29–41, 2009. (Cité page 1.)

S. Haber. The primate basal ganglia : parallel and integrative networks. *Journal of Chemical Neuroanatomy*, 26(3) :317–330, 2003. (Cité page 8.)

T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, et E.I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052) :801–806, Aug 2005. URL `http://dx.doi.org/10.1038/nature03721`. (Cité page 45.)

O. Hikosaka, Y. Takikawa, et R. Kawagoe. Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological reviews*, 80 (3) :953–978, 2000. (Cité pages 9 et 13.)

J. C. Houk, J. L. Adams, et A. G. Barto. A model of how the basal ganglia generate and use neural signals that predict reinforcement. Dans J. C. Houk, J. L. Davis, et D. G. Beiser, éditeurs, *Models of Information Processing in the Basal Ganglia*, pages 249–271. The MIT Press, Cambridge, MA, 1995. (Cité pages 8, 23 et 24.)

M. D. Humphries, K. Gurney, et T. J. Prescott. Is there a brainstem substrate for action selection ? *Philos Trans R Soc Lond B Biol Sci*, 362(1485) : 1627–1639, Sep 2007. URL `http://dx.doi.org/10.1098/rstb.2007.2057`. (Cité pages 13 et 60.)

M.D. Humphries, K. Gurney, et T.J. Prescott. Is There an Integrative Center in the Vertebrate Brain-Stem ? A Robotic Evaluation of a Model of the Reticular Formation Viewed as an Action Selection Device. *Adaptive Behavior*, 13(2) :97–113, 2005. (Cité pages 13 et 59.)

R.A. Jacobs, M.I. Jordan, S.J. Nowlan, et G.E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1) :79–87, 1991. (Cité pages 23, 24 et 25.)

D. Joel, Y. Niv, et E. Ruppin. Actor-critic models of the basal ganglia : new anatomical and computational perspectives. *Neural Networks*, 15(4–6), 2002. (Cité page 8.)

D. Joel et I. Weiner. The organization of the basal ganglia-thalamocortical circuits : open interconnected rather than closed segregated. *Neuroscience*, 63 :363–379, 1994. (Cité page 8.)

D. Joel et I. Weiner. The connections of the dopaminergic system with the striatum in rats and primates : An analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, 96(3) : 452–474, 2000. (Cité pages 8 et 31.)

T. Kanaseki et J.M. Sprague. Anatomical organization of the pretectal nuclei and tectal laminae in the cat. *Journal of Comparative Neurology*, 158 :319–338, 1974. (Cité page 9.)

C.R.S. Kaneko, C. Evinger, et A.F. Fuchs. Role of the cat pontine burst neurons in generation of saccadic eye movements. *J Neurophysiol*, 46(3) : 387–408, 1981. (Cité page 55.)

E.L. Keller. Participation of medial pontine reticular formation in eye movement generation in monkey. *J Neurophysiol*, 37(2) :316–332, 1974. (Cité page 55.)

A. E. Kelley. Neural integrative processes in the ventral striatum in relation to motivation, feeding and learning. Dans *INABIS'98*. 1998. (Cité page 31.)

A. E. Kelley. Neural integrative activities of nucleus accumbens subregions in relation to learning and motivation. *Psychobiology*, 27 :198–213, 1999. (Cité pages 8 et 31.)

Naomi Keren, Noam Peled, et Alon Korngreen. Constraining compartmental models using multiple voltage recordings and genetic algorithms. *J Neurophysiol*, 94(6) :3730–3742, Dec 2005. URL `http://dx.doi.org/10.1152/jn.00408.2005`. (Cité page 59.)

M. Khamassi. *Complementary roles of the rat prefrontal cortex and striatum in reward-based learning and shifting navigation strategies.* PhD thesis, Université Pierre et Marie Curie, Paris, 2007. (Cité page 37.)

M. Khamassi, B. Girard, A. Berthoz, et A. Guillot. Comparing three critic models of reinforcement learning in the basal ganglia connected to a detailed actor part in a S-R task. Dans F. Groen, N. Amato, A. Bonarini,

E. Yoshida, et B. Kröse, éditeurs, *Proceedings of the Eighth International Conference on Intelligent Autonomous Systems (IAS8)*, pages 430–437. IOS Press, Amsterdam, The Netherlands, 2004. (Cité pages 3 et 23.)

M. Khamassi, L. Lachèze, B. Girard, A. Berthoz, et A. Guillot. Actor-critic models of reinforcement learning in the basal ganglia : From natural to artificial rats. *Adaptive Behavior*, 13(2) :131–148, 2005. (Cité pages 3 et 23.)

M. Khamassi, L.E. Martinet, et A. Guillot. Combining self-organizing maps with mixtures of experts : Application to an actor-critic model of reinforcement learning in the basal ganglia. Dans S. Nolfi, G. Baldassare, R. Calabretta, J.C. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, et D. Parisi, éditeurs, *From Animals to Animats 9 : Proceedings of the ninth international conference on simulation of adaptive behavior,*, volume 4095 de *LNAI*, pages 394–405. Springer-Verlag, 2006. (Cité page 26.)

WL Kilmer, WS McCulloch, et J. Blum. A model of the vertebrate central command system. *International Journal of Man-Machine Studies*, 1 :279–309, 1969. (Cité page 60.)

J. J. Kim et M. G. Baxter. Multiple brain-memory systems : the whole does not equal the sum of its parts. *Trends Neurosci*, 24(6) :324–330, Jun 2001. (Cité page 37.)

W.M. King et F. Fuchs. Reticular control of vertical saccadic eye movements by mesencephalic burst neurons. *J Neurophysiol*, 42(3) :861–876, 1979. (Cité page 55.)

H. Kita, H. Tokuno, et A. Nambu. Monkey globus pallidus external segment neurons projecting to the neostriatum. *Neuroreport*, 10(7) :1476–1472, 1999. (Cité page 17.)

G.D. Konidaris et A. G. Barto. An adaptive robot motivational system. Dans S. Nolfi, G. Baldassarre, R. Calabretta, J.C.T. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, et D. Parisi, éditeurs, *From Animals to Animats 9 : Proceedings of the Ninth International Conference on Simulation of Adaptive Behavior*, LNCS. Springer, 2006. (Cité page 28.)

R.J. Krauzlis, D. Liston, et C.D. Carello. Target selection and the superior colliculus : goals, choices and hypotheses. *Vision Research*, 44 :1445–1451, 2004. (Cité page 11.)

J. Laurens et J. Droulez. Bayesian processing of vestibular information. *Biol Cybern*, 96(4) :389–404, Apr 2007. URL `http://dx.doi.org/10.1007/s00422-006-0133-1`. (Cité page 31.)

X. Li et M.A. Basso. Competitive stimulus interactions within single response fields of superior colliculus neurons. *The Journal of Neuroscience*, 25(49) :11357–11373, 2005. (Cité page 11.)

Xiaobing Li, Byounghoon Kim, et Michele A Basso. Transient pauses in delay-period activity of superior colliculus neurons. *J Neurophysiol*, 95 (4) :2252–2264, Apr 2006. URL `http://dx.doi.org/10.1152/jn.01000.2005`. (Cité page 11.)

J. Liénard, A. Guillot, et B. Girard. Multi-objective evolutionary algorithms to investigate neurocomputational issues : The case study of basal ganglia models. Dans *From Animals to Animats 11 : Proceedings of the Eleventh International Conference on Simulation of Adaptive Behavior*, LNAI, pages 602–609. Springer, 2010. (Cité pages 3 et 59.)

W. Lohmiller et J.J.E. Slotine. Contraction analysis for nonlinear systems. *Automatica*, 34(6) :683–696, 1998. (Cité pages 14 et 15.)

L. Manfredi, E. Maini, C. Laschi, P. Dario, B. Girard, N. Tabareau, et A. Berthoz. Implementation of a neurophysiologic model of saccadic movements on an anthropomorphic robotic head. Dans *IEEE-RAS Int. Conf. on Humanoid Robots*, pages 438–443, 2006. (Cité page 151.)

C. Masson et B. Girard. Decoding the grid cells for metric navigation using the residue numeral system. Dans *2nd International Conference on Cognitive Neurodynamics (ICCN2009)*, Hangzhou, China, 2009. (Cité pages 4 et 45.)

L.E. Mays et D.L. Sparks. Dissociation of visual and saccade-related responses in superior colliculus neurons. *J Neurophysiol*, 43(1) :207–232, 1980. (Cité pages 9 et 10.)

R. J. McDonald et N. M. White. A triple dissociation of memory systems : hippocampus, amygdala, and dorsal striatum. *Behav Neurosci*, 107(1) : 3–22, Feb 1993. (Cité page 37.)

J.G. McHaffie, T.R. Stanford, B.E. Stein, V. Coizet, et P. Redgrave. Subcortical loops through the basal ganglia. *Trends Neurosci*, 28(8) :401–407, Aug 2005. URL `http://dx.doi.org/10.1016/j.tins.2005.06.006`. (Cité pages 7, 10, 11 et 13.)

B.L. McNaughton, F.P. Battaglia, O. Jensen, E.I. Moser, et M.-B. Moser. Path integration and the neural basis of the 'cognitive map'. *Nat Rev Neurosci*, 7(8) :663–678, Aug 2006. URL `http://dx.doi.org/10.1038/nrn1932`. (Cité page 46.)

R.M. McPeek, J.H. Han, et E.L. Keller. Competition between saccade goals in the superior colliculus produces saccade curvature. *J Neurophysiol*, 89 (5) :2577–2590, May 2003. URL `http://dx.doi.org/10.1152/jn.00657.2002`. (Cité page 11.)

R.M. McPeek et E.L. Keller. Saccade target selection in the superior colliculus during a visual search task. *J Neurophysiol*, 88(4) :2019–2034, Oct 2002a. (Cité pages 10 et 11.)

R.M. McPeek et E.L. Keller. Superior colliculus activity related to concurrent processing of saccade goals in a visual search task. *J Neurophysiol*, 87(4) :1805–1815, Apr 2002b. URL `http://dx.doi.org/10.1152/jn.00501.2001`. (Cité page 11.)

R.M. McPeek et E.L. Keller. Deficits in saccade target selection after inactivation of superior colliculus. *Nat Neurosci*, 7(7) :757–763, Jul 2004. URL `http://dx.doi.org/10.1038/nn1269`. (Cité page 11.)

J.-A. Meyer et A. Guillot. Simulation of adaptive behavior in animats : review and prospect. Dans J. A. Meyer et S. W. Wilson, éditeurs, *From Animals to Animats : Proceedings of the First International Conference on the Simulation of Adaptive Behavior*. The MIT Press/Bradford Books, Cambridge, MA, 1991. (Cité page 1.)

J.-A. Meyer et A. Guillot. From SAB90 to SAB94 : Four years of animat research. Dans D. Cliff, P. Husbands, J.-A. Meyer, et S. W. Wilson, éditeurs, *From Animals to Animats 3 : Proceedings of the third international conference on simulation of adaptive behavior*. The MIT Press/Bradford Books, Cambridge, MA, 1994. (Cité page 1.)

J.-A. Meyer et A. Guillot. Handbook of robotics. Chapitre Biologically-inspired Robots, pages 1395–1422. Springer-Verlag, 2008. URL `http://animatlab.lip6.fr/papers/Draft_Biologically-inspiredf`. (Cité page 1.)

J.-A. Meyer, A. Guillot, B. Girard, M. Khamassi, P. Pirim, et A. Berthoz. The Psikharpax project : Towards building an artificial rat. *Robotics and autonomous systems*, 50(4) :211–223, 2005. (Cité page 157.)

F. A. Middleton et P. L. Strick. Basal ganglia and cerebellar loops : motor and cognitive circuits. *Brain Res Brain Res Rev*, 31(2-3) :236–250, Mar 2000. (Cité page 7.)

J. W. Mink. The basal ganglia : Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50(4) :381–425, 1996. (Cité pages 8 et 13.)

A.K. Moschovakis. The superior colliculus and eye movement control. *Current Opinion in Neurobiology*, 6 :811–816, 1996. (Cité page 9.)

A.K. Moschovakis, T. Kitama, Y. Dalezios, J. Petit, A.M. Brandi, et A.A. Grantyn. An anatomical substrate for the spatiotemporal transformation. *J Neurosci*, 18(23) :10219–10229, 1998. (Cité pages 55 et 56.)

A.K. Moschovakis, C.A. Scudder, et S.M. Highstein. The microscopic anatomy and physiology of the mammalian saccadic system. *Prog Neurobiol*, 50 :133–254, 1996. (Cité page 9.)

E.I. Moser, E. Kropff, et M.-B. Moser. Place cells, grid cells, and the brain's spatial representation system. *Annu Rev Neurosci*, 31 :69–89, 2008. URL `http://dx.doi.org/10.1146/annurev.neuro.31.061307.090723`. (Cité page 46.)

J.-B. Mouret, S. Doncieux, et B. Girard. Importing the computational neuroscience toolbox into neuro-evolution—application to basal ganglia. Dans *Genetic and Evolutionary Computation Conference 2010 (GECCO2010)*, 2010. (Cité page 60.)

D. P. Munoz et R. H. Wurtz. Saccade-related activity in monkey superior colliculus. i. characteristics of burst and buildup cells. *J Neurophysiol*, 73 (6) :2313–2333, Jun 1995. (Cité page 10.)

D.P. Munoz, D.M. Waitzman, et R.H. Wurtz. Activity of neurons in monkey superior colliculus during interrupted saccades. *Journal of Neurphysiology*, 75(6) :2562–2580, 1996. (Cité page 55.)

S. F. Neggers et H. Bekkering. Ocular gaze is anchored to the target of an ongoing pointing movement. *J Neurophysiol*, 83(2) :639–651, Feb 2000. (Cité page 56.)

S. F. Neggers et H. Bekkering. Coordinated control of eye and hand movements in dynamic reaching. *Hum Mov Sci*, 21(3) :349–376, Sep 2002. (Cité page 56.)

S. N'Guyen, P. Pirim, J.-A. Meyer, et B. Girard. An integrated neuromimetic model of the saccadic eye movements for the psikharpax robot. Dans *From Animals to Animats 11 : Proceedings of the Eleventh International Conference on Simulation of Adaptive Behavior*, LNAI, pages 115–126. Springer, 2010. (Cité pages 3, 56 et 58.)

Y. Niv, N.D. Daw, D. Joel, et P. Dayan. Tonic dopamine : opportunity costs and the control of response vigor. *Psychopharmacology (Berl)*, 191(3) :507–520, Apr 2007. URL `http://dx.doi.org/10.1007/s00213-006-0502-4`. (Cité page 8.)

M. G. Packard, R. Hirsh, et N. M. White. Differential effects of fornix and caudate nucleus lesions on two radial maze tasks : evidence for multiple memory systems. *J Neurosci*, 9(5) :1465–1472, May 1989. (Cité page 37.)

A. Parent, F. Sato, Y. Wu, J. Gauthier, M. Lévesque, et M. Parent. Organization of the basal ganglia : the importance of the axonal collateralization. *Trends in Neuroscience*, 23(10) :S20–S27, 2000. (Cité page 17.)

C. Parron et E. Save. Evidence for entorhinal and parietal cortices involvement in path integration in the rat. *Exp Brain Res*, 159(3) :349–359, Dec 2004. URL `http://dx.doi.org/10.1007/s00221-004-1960-8`. (Cité page 46.)

J.M. Pearce, A.D. Roberts, et M. Good. Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. *Nature*, 396 (6706) :75–77, 1998. (Cité pages 40, 41 et 44.)

T. J. Prescott, F. Montes-Gonzalez, K. Gurney, M. D. Humphries, et P. Redgrave. A robot model of the basal ganglia : Behavior and intrinsic processing. *Neural Networks*, 19 :31–61, 2006. (Cité pages 14, 18, 19 et 22.)

T. J. Prescott, P. Redgrave, et K. N. Gurney. Layered control architectures in robot and vertebrates. *Adaptive Behavior*, 7(1) :99–127, 1999. (Cité page 13.)

T.J. Prescott, J.J. Bryson, et A.K. Seth. Introduction. modelling natural action selection. *Philos Trans R Soc Lond B Biol Sci*, 362(1485) :1521–1529, Sep 2007. URL `http://dx.doi.org/10.1098/rstb.2007.2050`. (Cité page 13.)

Z. W. Pylyshyn et R. W. Storm. Tracking multiple independent targets : evidence for a parallel tracking mechanism. *Spat Vis*, 3(3) :179–197, 1988. (Cité page 31.)

Rajesh P N Rao. Bayesian computation in recurrent neural circuits. *Neural Comput*, 16(1) :1–38, Jan 2004. (Cité page 35.)

P. Redgrave. Basal ganglia. *Scholarpedia*, 2(6) :1825, 2007. (Cité pages 5 et 7.)

P. Redgrave, T. J. Prescott, et K. Gurney. The basal ganglia : a vertebrate solution to the selection problem ? *Neuroscience*, 89(4) :1009–1023, 1999. (Cité pages 8 et 13.)

A.D. Redish. *Beyond the cognitive map : From place cells to episodic memory*. MIT Press, 1999. (Cité page 37.)

L. Rondi-Reig, G.H. Petit, C. Tobin, S. Tonegawa, J. Mariani, et A. Berthoz. 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, Apr 2006. URL `http://dx.doi.org/10.1523/JNEUROSCI.3408-05.2006`. (Cité page 49.)

D. Samu, P. Eros, B. Ujfalussy, et T. Kiss. Robust path integration in the entorhinal grid cell system with hippocampal feed-back. *Biol Cybern*, 101(1) :19–34, Jul 2009. URL `http://dx.doi.org/10.1007/s00422-009-0311-z`. (Cité page 48.)

F. Sato, P. Lavallee, M. Lévesque, et A. Parent. Single-axon tracing study of neurons of the external segment of the globus pallidus in primates. *Journal of Comparative Neurology*, 417 :17–31, 2000. (Cité page 17.)

S. Schaal, Y. Nakamura, et P. Dario. Robotics and Neuroscience special issue, 21(4). *Neural Networks*, 2008. (Cité page 2.)

W. Schultz, P. Dayan, et P. R. Montague. A neural substrate of prediction and reward. *Science*, 275 :1593–1599, 1997. (Cité page 8.)

N. Schweighofer, MA Arbib, et PF Dominey. A model of the cerebellum in adaptive control of saccadic gain. II. Simulation results. *Biological Cybernetics*, 75(1) :29–36, 1996a. (Cité page 58.)

N. Schweighofer, M.A. Arbib, et P.F. Dominey. A model of the cerebellum in adaptive control of saccadic gain. I. The model and its biological substrate. *Biological Cybernetics*, 75(1) :19–28, 1996b. (Cité page 58.)

C.A. Scudder, C.R.S. Kaneko, et A.F. Fuchs. The brainstem burst generator for saccadic eye movements. A modern synthesis. *Exp Brain Res*, 142 : 439–462, 2002. (Cité pages 9 et 55.)

T. J. Sejnowski, C. Koch, et P. S. Churchland. Computational neuroscience. *Science*, 241(4871) :1299–1306, Sep 1988. (Cité page 1.)

D. Sheynikhovich. *Spatial navigation in geometric mazes : a computational model of rodent behavior*. PhD thesis, EPFL, Lausanne, CH, 2007. (Cité page 47.)

R. Soetedjo, C.R.S. Kaneko, et A.F. Fuchs. Evidence that the superior colliculus participates in the feedback control of saccadic eye movements. *Journal of Neurophysiology*, 87 :679–695, 2000. (Cité page 55.)

T. Solstad, C.N. Boccara, E. Kropff, M.-B. Moser, et E.I. Moser. Representation of geometric borders in the entorhinal cortex. *Science*, 322 (5909) :1865–1868, Dec 2008. URL `http://dx.doi.org/10.1126/science.1166466`. (Cité page 48.)

W. Staines, S. Atmadja, et H. Fibiger. Demonstration of a pallidostriatal pathway by retrograde transport of HRP-labelled lectin. *Brain Research*, 206 :446–450, 1981. (Cité page 17.)

B.E. Stein et M.A. Meredith. *The Merging of the Senses*. MIT Press, 1993. (Cité page 11.)

H. Sun et T. Yao. A neural-like approach to residue-to-decimal conversion. Dans *IEEE International Conference on Neural Networks*, volume 6, pages 3883–3887, 1994. (Cité page 46.)

R. E. Suri et W. Schultz. Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Exp Brain Res*, 121(3) :350–354, Aug 1998. (Cité page 24.)

N. Tabareau, D. Bennequin, A. Berthoz, J.-J. Slotine, et B. Girard. Geometry of the superior colliculus mapping and efficient oculomotor computation. *Biological Cybernetics*, 97(4) :279–292, 2007. (Cité pages 4, 51, 52, 55, 58 et 151.)

N. Tabareau et J.J. Slotine. Notes on Contraction Theory. *Arxiv preprint nlin.AO/0601011*, 2006. (Cité page 15.)

T. Tanner, L. Canto-Pereira, et H. Buelthoff. Free vs. constrained gaze in a multiple-object-tracking paradigm. Dans *30th European Conference on Visual Perception*, Arezzo, Italy, 2007. (Cité page 34.)

J.M. Tepper et J.P. Bolam. Functional density and specificity of neostrital interneurons. *Current Opinion in Neurobiology*, 14 :685–692, 2004. (Cité page 16.)

J.M. Tepper, T. Koós, et C.J. Wilson. Gabaergic microcircuits in the neostriatum. *Trends in Neuroscience*, 11 :662–669, 2004. (Cité page 16.)

M.-T. Tran, P. Souères, M. Taïx, et B. Girard. A computational approach from robotics for testing eye-centered vs body-centered reaching control. Dans *Progress in Motor Control 2009*, 2009a. (Cité pages 4 et 57.)

M.-T. Tran, Ph. Souères, M. Taïx, et B. Girard. Eye-centered vs. body-centered reaching control : A robotics insight into the neuroscience debate. Dans *IEEE International Conference on Robotics and Biomimetics (ROBIO 2009)*, 2009b. (Cité pages 4 et 57.)

O. Trullier, S. Wiener, A. Berthoz, et J.-A. Meyer. Biologically-based artificial navigation systems : Review and prospects. *Progress in Neurobiology*, 51 :483–544, 1997. (Cité pages 44 et 48.)

B. Ujfalussy, P. Eros, Z. Somogyvari, et T. Kiss. Episodes in space : A modelling study of hippocampal place representation. Dans M. Asada, J.C.T. Hallam, J.-A. Meyer, et J. Tani, éditeurs, *From Animals to Animats 10 : Proceedings of the Tenth International Conference on Simulation of Adaptive Behavior*, volume 5040 de *LNAI*, pages 123–136. Springer, 2008. (Cité page 39.)

J.A. van Gisbergen, A.J. van Opstal, et A.A. Tax. Collicular ensemble coding of saccades based on vector summation. *Neuroscience*, 21(2) :541–555, May 1987. (Cité pages 52, 54 et 55.)

A. J. van Opstal et H. H L M Goossens. Linear ensemble-coding in midbrain superior colliculus specifies the saccade kinematics. *Biol Cybern*, 98(6) :561–577, Jun 2008. URL `http://dx.doi.org/10.1007/s00422-008-0219-z`. (Cité pages 55 et 56.)

B. Webb et T. R. Consi, éditeurs. *Biorobotics, methods and applications*. AAAI Press/MITPress, Cambridge, MA, 2001. (Cité page 1.)

Q. Wei, S. Sueda, et D.K. Pai. Biomechanical simulation of human eye movement. Dans *The 5th International Symposium on Biomedical Simulation ISBMS10*, 2010. (Cité page 58.)

W. Werner. Neurons in the primate superior colliculus are active before and during arm movements to visual targets. *Eur J Neurosci*, 5(4) :335–340, Apr 1993. (Cité page 9.)

W. Werner, S. Dannenberg, et K. P. Hoffmann. Arm-movement-related neurons in the primate superior colliculus and underlying reticular formation : comparison of neuronal activity with emgs of muscles of the shoulder, arm and trunk during reaching. *Exp Brain Res*, 115(2) :191–205, Jun 1997a. (Cité pages 9 et 11.)

W. Werner, K. P. Hoffmann, et S. Dannenberg. Anatomical distribution of arm-movement-related neurons in the primate superior colliculus and underlying reticular formation in comparison with visual and saccadic cells. *Exp Brain Res*, 115(2) :206–216, Jun 1997b. (Cité pages 9 et 11.)

N.M. White et R.J. McDonald. Multiple parallel memory systems in the brain of the rat. *Neurobiol Learn Mem*, 77(2) :125–184, Mar 2002. URL `http://dx.doi.org/10.1006/nlme.2001.4008`. (Cité page 37.)

K. Yoshida, R. McCrea, A. Berthoz, et P.P. Vidal. Morphological and physiological characteristics of inhibitory burst neurons controlling horizontal rapid eye movements in the alert cat. *J Neurophysiol*, 48(3) :761–784, 1982. (Cité page 56.)

D. Zhang et A. Nagurney. On the stability of projected dynamical systems. *Journal of Optimization Theory and Applications*, 85(1) :97–124, 1995. (Cité page 15.)

Acronymes

AE	Algorithmes évolutionnistes
BG	Ganglions de la base
BN	Neurones oculomoteurs du colliculus supérieur
BUN	Neurones oculomoteurs avec activité soutenue du colliculus supérieur
CBG	Modèle contractant des ganglions de la base (Girard et al., 2008)
CBTC	Circuits cortico-baso-thalamo-corticaux
cMRF	Région centrale de la formation réticulée médiale
CRT	Théorème Chinois des restes
BG	Ganglions de la base
BN	Neurones oculomoteurs du colliculus supérieur
BUN	Neurones oculomoteurs avec activité soutenue du colliculus supérieur
DLS	Striatum dorso-latéral
dMEC	Bande dorso-latérale du cortex entorhinal médian
DMS	Striatum dorso médian
EBN	Neurones excitateurs à bouffées d'activité des générateurs de saccades
FEF	Champs oculaires frontaux
FOR	Région fastigiale oculomotrice
GC	Cellules de grille
GPe	Globus pallidus externe
GPi	Globus pallidus interne
GPR	Modèle des ganglions de la base de Gurney et al. (2001a,b)
KS	test statistique de Kolmogorov-Smirnoff
LIP	Cortex latéral intra-pariétal
lPDS	Systèmes dynamiques projetés localement
M	Neurones du colliculus supérieur ayant une activité maintenue entre la
MOT	Tâche de suivi d'objets multiples
	disparition du stimulus et le mouvement vers ce stimulus.
mRF	Formation Réticulée médiale
NAcc	Noyau accumbens
NRTP	Noyau réticulé tegmental du pont
PPN	Noyau pédonculopontin
PRTR	Stratégie de navigation d'action déclenchée par la reconnaissance d'un lieu
QV	Neurones quasi-visuels du colliculus supérieur
R	Neurones moteurs d'atteinte par le bras (*reach*) du colliculus supérieur
RNS	Système numéral à base de restes
S	Neurones moteurs saccadiques du colliculus supérieur
SBG	Générateurs de saccades du tronc cérébral

SC	Colliculus supérieur
SEF	Champs oculaires supplémentaires
SNc	Substance noire compacte
SNr	Substance noire réticulée
S-R	Association stimulus-réponse
STN	Noyau subthalamique
STT	Transformation spatio-temporelle
TH	Un quelconque noyau thalamus impliqués dans les boucles CBTC
TRN	Noyau thalamique réticulé
V	Cellules visuelles du colliculus supérieur
WTA	Processus de sélection « winner-takes-all »

Ce document a été préparé à l'aide de l'éditeur de texte GNU Emacs, du logiciel de dessin vectoriel Xfig, du logiciel de visualisation scientifique gnuplot, d'une quantité indécente de caféine et du logiciel de composition typographique LaTeX 2_ε.

Titre Modélisation neuromimétique : Sélection de l'action, navigation et exécution motrice

Résumé Ce mémoire d'Habilitation à Diriger des Recherches synthétise les travaux que j'ai menés dans le domaine des neurosciences computationnelles. Ils traitent de trois thématiques principales en interaction : la sélection de l'action, la navigation et l'exécution motrice. Le substrat neural de ces fonctions, et principalement les ganglions de la base et le colliculus supérieur, ont été modélisés sous forme de réseaux de neurones contraints par les données issues de la neuro-anatomie et de l'électrophysiologie. Les résultats présentés résument mes contributions portant sur : les processus de sélection, d'apprentissage par renforcement et de modulation motivationnelle dans les ganglions de la base, le rôle de l'incertitude dans la sélection de l'action, la sélection de stratégies de navigation, l'intégration de chemin pour la stratégie de retour au point de départ, et la transformation spatio-temporelle pour la génération de saccades oculaires. Enfin, les liens reliant l'ensemble de ces études sont mis en exergue afin de délimiter le programme de recherche qui en découle.

Mots-clés Neurosciences computationnelles, neuro-robotique, sélection de l'action, navigation, exécution motrice, ganglions de la base, colliculus supérieur

Title Neuromimetic Models for Action Selection, Navigation and Motor Execution

Abstract This thesis synthesizes the studies I have carried out in the domain of computational neuroscience. They deal with three interacting main topics : action selection, navigation and motor execution. The neural substrate of these functions, and particularly the basal ganglia and the superior colliculus, have been modeled using neural networks constrained by neuroanatomical and electrophysiological data. The presented results summarize my contributions to : selection, reinforcement learning and motivational modulation processes in the basal ganglia, the role of uncertainty in action selection, the selection of navigation strategies, path integration for the homing strategy, and the spatio-temporal transformation for the generation of ocular saccades. Finally, the links between these studies is emphasized so as to define the resulting research program.

Keywords Computational neuroscience, neuro-robotics, action selection, navigation, motor execution, basal ganglia, superior colliculus