



**HAL**  
open science

# Auto-organisation multi-échelle pour l'émergence de comportements sensorimoteurs coordonnés

Tristan Gillard

► **To cite this version:**

Tristan Gillard. Auto-organisation multi-échelle pour l'émergence de comportements sensorimoteurs coordonnés. Informatique [cs]. Université de Lorraine, 2022. Français. NNT: 2022LORR0353 . tel-04105974

**HAL Id: tel-04105974**

**<https://hal.univ-lorraine.fr/tel-04105974v1>**

Submitted on 25 May 2023

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

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



**UNIVERSITÉ  
DE LORRAINE**

**BIBLIOTHÈQUES  
UNIVERSITAIRES**

## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)  
*(Cette adresse ne permet pas de contacter les auteurs)*

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

---

# Auto-organisation multi-échelle pour l'émergence de comportements sensorimoteurs coordonnés

## THÈSE

présentée et soutenue publiquement le 12 décembre 2022

pour l'obtention du

**Doctorat de l'Université de Lorraine**  
(mention informatique)

par

Tristan GILLARD

### Composition du jury

<i>Président, rapporteur :</i>	Nicolas BRÉDÈCHE	Professeur
<i>Rapporteur :</i>	Emmanuel DAUCÉ	Maître de Conférences habilité à diriger des recherches
<i>Examinatrices :</i>	Anne BOYER Audrey DUSSUTOUR	Professeur Directrice de Recherche au CNRS
<i>Directeurs de thèse :</i>	Alain DUTECH Jérémy FIX	Chargé de Recherche Inria habilité à diriger des recherches Maître de Conférences

Mis en page avec la classe thesul.



## Remerciements

Noire! Voilà déjà de quoi contrer cette page blanche histoire que je ne reste pas devant indéfiniment. Bien que la thèse soit maintenant rédigée, cet exercice des remerciements demeure et n'est pas particulièrement le plus simple. L'expression n'a jamais été mon point fort, et pourtant il s'agit bien ici que j'arrive à exprimer tous mes sentiments de gratitude envers ceux qui m'ont aidé, d'une façon ou d'une autre, pour l'élaboration de ces travaux. Je ne me fais cependant pas d'illusions, mes mots ne seront à la hauteur ni de ce que je ressens ni de vos diverses contributions.

Tout d'abord, je tiens à remercier les différents membres du jury pour le temps qu'ils ont consacré à ma thèse. Merci à Nicolas BRÉDÈCHE et Emmanuel DAUCÉ pour leurs relectures et leurs rapports, ainsi que pour les discussions lors du pot de thèse. Merci à Anne BOYER et à Audrey DUSSUTOUR pour leur rôle d'examinatrices. En particulier merci Audrey d'avoir fait le "voyage" depuis le domaine de la biologie pour faire partie de mon jury et pour les compliments sur ma partie biologie, cela m'a vraiment touché!

Je tiens aussi à remercier mes anciens encadrants de stage sans lesquels je n'en serais pas là aujourd'hui. Merci à Sophie PINCHINAT, Thomas GENET, Jean LIEBER et à Emmanuel NAUER pour leur enthousiasme, leur énergie, leurs conseils et leur bienveillance. Merci de m'avoir accompagné lors de mes premiers pas dans le monde de la recherche.

Merci beaucoup à mes directeurs de thèse, Alain DUTECH et Jérémy FIX, pour m'avoir fait confiance et m'avoir suivi tout au long de mes recherches. Je suis très heureux d'avoir pu faire ces travaux avec vous. Votre bienveillance et votre soutien permanent, y compris pendant le covid, m'ont été des atouts indispensables. Encore merci à vous!

Un énorme merci aux équipes pédagogiques de l'IUT Nancy Charlemagne avec qui j'ai travaillé au cours de ma thèse. Merci aussi pour l'opportunité d'avoir siégé au conseil restreint de l'IUT. En particulier, je remercie Nadia BELLALEM pour avoir réussi à me trouver un poste dès la première année et pour m'avoir accepté dans son équipe à de multiples reprises. Je remercie aussi Vincent THOMAS pour nos échanges et ses conseils sur la pédagogie, ainsi que pour ses encouragements rassurants. Enfin, merci à Blandine GALTIE et à Françoise KERN pour leur bonne humeur et pour m'avoir simplifié la vie à de multiples reprises vis-à-vis de l'emploi du temps.

Un merci sucré pour les autres "Biscuits" : Bernard GIRAU, Yann BONNIFACE et Hervé FREZZA-BUET, pour leurs retours critiques, parfois salés, mais toujours bienvenus! L'escapade dans les Vosges a été un super moment, autant d'échanges scientifiques que de détente, et reste l'un de mes meilleurs souvenirs de la thèse. Je remercie Antoinette COURRIER pour son accompagnement dans les démarches administratives diverses, ce fut nettement moins pénible grâce à vous.

Merci à l'équipe des petits "Biscuits", Adrien, Yann et Noémie, ainsi qu'à Pierre, pour les diverses discussions sérieuses (...ou moins!) pendant les pauses. Et bon courage à Noémie et à Pierre pour leurs soutenances qui arrivent maintenant à grands pas!

Je finirai ces remerciements par ceux concernant mes proches, même s'ils sont pour nombre d'entre eux à l'autre bout de la France voire à l'autre bout du monde!

Déjà, merci à Théo, pour les pauses restos sur Metz et les longues discussions captivantes qui

s'ensuivaient.

Ensuite, un grand merci à mes chers "Erwendaliens", piliers garants de ma santé bonne humeur bien avant que je commence ma thèse et très certainement bien après que je l'eusse terminée. Merci pour les soirées (déjantées ou non!), les franches rigolades toujours, et ces amitiés sincères que nous avons réussi à conserver contre vents et marées depuis toutes ces années! Merci à Ama, Blak, Charlène, Doc, Elea, Émilie, Honpah, Morin, Mini, Mizu, Shikaa, Titouan.

Côté famille, je remercie mes beaux parents Yvon et Isabelle pour leur gentillesse, les invitations les dimanches, et les dépannages de machine à laver ou de chaudière en plein hiver, ainsi que Michel pour nous avoir hébergé gracieusement avec ma femme tout au long de ma thèse et au delà. Bien évidemment, je remercie énormément mes parents, Maël et Sylvie, pour leur amour inconditionnel et l'éducation qu'ils m'ont apportée. La rigueur et la curiosité nécessaires pour la recherche, la remise en question de ce que j'entends ou lis, c'est initialement à vous que je le dois, à travers nos nombreuses discussions familiales et diverses argumentations. Merci beaucoup, et je ne le dirai jamais assez : je vous aime! Une mention spéciale à mon grand frère Florian. Merci pour toutes ces années de complicité fraternelle. Sans cela, dur de savoir si j'aurais trouvé la motivation de retourner sur les bancs de la fac pour me réorienter en informatique.

Merci à Osiris, l'adorable petite boule de poil toujours à squatter mes genoux pour un brin de chaleur supplémentaire.

Enfin, un très grand merci à toi, Audrey, pour ton amour, ta patience et ton soutien à toute épreuve. Avec la fin de cette thèse, c'est un sacré bout d'aventure qui s'achève, mais la suite s'annonce tout autant passionnante et mouvementée!

*À Piwi, que j'ai vraiment hâte de rencontrer !*



# Table des matières

<b>Chapitre 1 Introduction</b>	<b>1</b>
<b>Chapitre 2 État de l'art</b>	<b>5</b>
2.1 Biologie . . . . .	5
2.1.1 Comportements . . . . .	5
2.1.2 Apprentissages . . . . .	7
2.2 Agents artificiels utilisant de l'apprentissage non-associatif . . . . .	21
2.2.1 Modèles conceptuels . . . . .	21
2.2.2 Utilisation de l'habituation pour la détection de nouveauté . . . . .	23
2.2.3 Autres travaux opérationnels . . . . .	24
2.3 Bilan de l'état de l'art . . . . .	25
<b>Chapitre 3 IDSM</b>	<b>27</b>
3.1 Modèle . . . . .	28
3.2 Exemples d'IDSM simples . . . . .	32
3.2.1 Cercle . . . . .	32
3.2.2 Agent avec deux moteurs et deux senseurs . . . . .	33
3.3 Exemples de combinaisons de comportements . . . . .	40
3.3.1 Entraînements successifs . . . . .	44
3.3.2 Entraînements séparés . . . . .	46
3.3.3 Comportements dépendants d'un senseur . . . . .	46
3.3.4 Bilan de ces expériences . . . . .	52
3.4 Effets des paramètres de l'IDSM et autres points d'attention . . . . .	56
3.4.1 Un seul nœud par position sensorimotrice . . . . .	56
3.4.2 Le poids des nœuds $N_w$ . . . . .	56
3.4.3 Hyperparamètre $k_\omega$ pour la fonction de poids $\omega(N_w)$ . . . . .	58
3.4.4 Hyperparamètre $k_d$ pour la fonction de distance $d(\mathbf{N}_p, \mathbf{x}_{SM})$ . . . . .	58
3.4.5 Le paramètre de délai d'activation des nœuds $\tau$ . . . . .	60

3.4.6	Délai minimum $\Delta t_{nodes}$ entre deux créations de noeud . . . . .	62
3.4.7	Seuil de saturation de nœuds $k_t$ . . . . .	63
3.4.8	Commande motrice $\mu$ . . . . .	65
3.4.9	Calcul de la variation de la commande motrice $\dot{\mu}(t)$ . . . . .	66
3.5	Motivations . . . . .	66
<b>Chapitre 4 IDSM avec apprentissages non-associatifs</b>		<b>69</b>
4.1	Habitude . . . . .	69
4.1.1	Modèles . . . . .	71
4.1.2	Expériences et résultats . . . . .	74
4.1.3	Discussion autour de nos modèles d'habitude . . . . .	86
4.1.4	Bilan sur l'habitude . . . . .	98
4.2	Sensibilisation spécifique au site . . . . .	98
4.2.1	Modèles . . . . .	100
4.2.2	Expériences et résultats . . . . .	101
4.2.3	Bilan sur la sensibilisation spécifique au site . . . . .	112
4.3	Pseudo-conditionnement . . . . .	114
4.3.1	Modèle avec masque sur les nœuds . . . . .	115
4.3.2	Expériences et résultats . . . . .	116
4.3.3	Bilan sur le pseudo-conditionnement . . . . .	120
4.4	Bilan de ces modèles . . . . .	121
<b>Chapitre 5 Bilan et Perspectives</b>		<b>127</b>
<b>Bibliographie</b>		<b>135</b>

# Table des figures

2.1	Cycle de vie de <i>Physarum polycephalum</i> (figure adaptée de BARRANTES, LEIPZIG et MARWAN, 2012 et corrigée en suivant HARA et OHTA, 1981). Ses deux phases mobiles sont l'amibe (avec un noyau haploïde) et le plasmode (avec plusieurs noyaux diploïdes). La fusion de deux amibes (syngamie) forme un zygote qui donne un plasmode après division de son noyau. En cas de manque de nutriments et de présence de lumière, le plasmode va se couvrir de sporanges. Arrivées à maturation, ces sporanges relâchent les spores, qui libèrent les amibes après germination. . . .	13
2.2	Effet des sensibilisations spécifiques au site court terme et long terme (figure 6 de GROVES, LEE et THOMPSON, 1969), avec une fréquence de 16 stimuli par seconde et avec une intensité de 10V. La sensibilisation spécifique au site à court terme est relativement éphémère, le pic est atteint en 5 réponses et la réponse la plus haute vaut 140% de la réponse témoin. L'effet s'estompe vers la 15 <sup>e</sup> réponse. La sensibilisation spécifique au site à long terme est plus lente et progressive à se mettre en place, elle commence à la 25 <sup>e</sup> réponse et atteint un niveau asymptotique d'environ 160% de la réponse témoin après la 120 <sup>e</sup> réponse. . . . .	15
2.3	Effets de l'intensité et de la fréquence du stimulus sur l'habituation et la sensibilisation spécifique au site (figure 7 de GROVES, LEE et THOMPSON, 1969). L'habituation est prédominante à faible intensité, et est plus marquée proportionnellement à la fréquence. La sensibilisation spécifique au site est, à l'inverse, prédominante à forte intensité et à basse fréquence. . . . .	16
3.1	L'IDSM est basé sur l'espace sensorimoteur normalisé de l'agent. À partir de la position sensorimotrice actuelle de l'agent $\mathbf{x}_{SM}$ dans cet espace et des différents nœuds de l'IDSM, la commande motrice $\boldsymbol{\mu}$ est calculée. Cette commande contrôle le comportement de l'agent dans son environnement. $S_E$ est l'état complet de l'environnement. . . . .	27
3.2	Un IDSM est construit à partir d'un ensemble de nœuds $N^i$ qui influencent la dynamique de l'agent dans l'espace sensorimoteur SM à travers des composantes vitesse $\mathbf{V}^i$ et attraction $\mathbf{A}^i$ . . . . .	28
3.3	Fonctions non linéaires utilisées pour calculer la densité de nœuds d'un état sensorimoteur, et pour mettre à l'échelle l'influence des nœuds en fonction de leur proximité avec l'état sensorimoteur actuel (à gauche), selon l'équation 3.4, et de leur poids (à droite), selon l'équation 3.3. . . . .	30

3.4	Affichage de la densité $\phi_{\mathcal{N}}$ pour un IDSM en une seule dimension. Le graphique de gauche montre l'impact de la variation du poids du nœud $N^2$ en $\mathbf{N}_{\mathbf{p}}^2 = [0.9]$ . Son poids $N_w^2$ vaut $-2000$ (en bleu), $-100$ (orange), $0$ (vert), $100$ (rouge) et $2000$ (violet). Le poids du premier nœud $N^1$ en $\mathbf{N}_{\mathbf{p}}^1 = [0.25]$ reste à $0$ . Le second graphique donne la densité pour 20 nœuds à positions et poids aléatoires (position dans $[0, 1]$ , poids dans $[-100, 0]$ ) pour donner un exemple <i>théorique</i> d'un paysage de densité $\phi_{\mathcal{N}}$ . Les lignes verticales rouges représentent la position des nœuds, la ligne horizontale verte indique le seuil par défaut $k_t = 1$ en dessous duquel la création de nœud serait possible. Dans cet exemple aléatoire, la création de nœud n'est pas possible dans la majorité de l'espace. . . . .	30
3.5	Résultats d'un IDSM permettant à un agent doté de deux moteurs de tourner en rond. Le premier graphique de 3.5a montre la dynamique de l'espace moteur en fin d'entraînement et la trajectoire que l'agent a suivi. Le second graphique indique les différents nœuds créés pendant l'entraînement, la base de la flèche étant le $\mathbf{N}_{\mathbf{p}}$ , la flèche le $\mathbf{N}_{\mathbf{v}}$ et l'intensité de la couleur le poids du nœud relatif aux autres. Le premier graphique de 3.5b (resp. 3.5c) montre la dynamique en fin d'expérience lorsque la composante motrice $[\mathbf{x}_{SM}]^{\mu}$ de l'agent a été repositionné en $(0.13, 0.17)$ (resp. $(0.51, 0.5)$ ) avant de laisser le contrôle à l'IDSM, avec les trajectoires de l'agent dans l'espace moteur. Les seconds graphiques de 3.5b et 3.5c montrent les trajectoires de l'agent dans l'environnement pendant l'entraînement et suite au repositionnement dans l'espace moteur. Dans les deux cas, l'agent reprend rapidement une trajectoire identique à celle d'entraînement. . . . .	34
3.6	Robot avec deux roues et deux senseurs de lumière. . . . .	35
3.7	Boîtes à moustache pour comparer les différences de distances entre l'agent et la lampe entre chaque trajectoire après relocalisation de l'agent pour le comportement phototaxie. Pour chaque trajectoire, la boîte est calculée à partir des 250 distances, une par pas de temps. Pour chaque boîte, le trait orange plein indique la médiane, le trait vert pointillé indique la moyenne, la croix bleu est la distance depuis la dernière position de l'agent. Les quatre premières trajectoires sont celles d'entraînement avec marche forcée. Pour toutes les trajectoires, l'agent se stabilise entre 1.3 et 1.4 unité de distance de la lampe. . . . .	36
3.8	Quelques exemples de trajectoires de l'expérience sur le comportement phototaxie. Dans tous les cas, que ce soit pendant l'entraînement ou avec l'IDSM, l'agent arrive à se stabiliser face à la lampe à une distance entre 1.3 et 1.4 d'elle. . . . .	37
3.9	Boîtes à moustache pour comparer les différences de distances entre l'agent et la lampe entre chaque trajectoire après relocalisation de l'agent pour le comportement phototaxie. Pour chaque trajectoire, la boîte est calculée à partir des 250 distances, une par pas de temps. Pour chaque boîte, le trait orange plein indique la médiane, le trait vert pointillé indique la moyenne, la croix bleu est la distance depuis la dernière position de l'agent. Les quatre premières trajectoires sont celles d'entraînement avec marche forcée. Contrairement aux résultats du comportement phototaxie, les distances sont très variées à la fois pour une même trajectoire mais aussi entre trajectoires. . . . .	38
3.10	Exemples de trajectoires de l'expérience sur le comportement photophobie illustrant les trois comportements de fuite observés (en fonction de la position de l'agent et de son état moteur de départ) : oscillations horizontales, oscillations verticales, ou mixte des deux le temps de se stabiliser. . . . .	39



---

3.11	Boîtes à moustache pour comparer les différences de distances entre l'agent et la lampe entre chaque trajectoire après relocalisation de l'agent pour le comportement phototaxie. Pour chaque trajectoire, la boîte est calculée à partir des 250 distances, une par pas de temps. Pour chaque boîte, le trait orange plein indique la médiane, le trait vert pointillé indique la moyenne, la croix bleu est la distance depuis la dernière position de l'agent. Les quatre premières trajectoires sont celles d'entraînement avec marche forcée. Le graphique permet de faire ressortir deux comportements : le premier qui permet à l'agent d'aller se stabiliser à une distance entre 1.15 et 1.2 de la lampe, et le second où il reste plus loin de la lampe, <i>a priori</i> sans stabiliser car les boîtes ressemblent plus à celles de photophobie (larges et avec la distance de la dernière position loin de la médiane) qu'à celle de phototaxie (petites et avec la dernière distance très proche de la médiane). Ce second comportement n'a pas été appris pendant l'entraînement et est l'agent qui fait des ronds plutôt que de se rapprocher de la lampe. . . . .	40
3.12	Exemples de trajectoires de l'expérience sur le comportement phototaxie sinusoïdale. 3.12a et 3.12b montrent le comportement attendu de l'agent : serpenter vers la lampe et se stabiliser à une certaine distance d'elle (entre 1.15 et 1.2). Par contre, 3.12c montre un comportement émergent de la dynamique des nœuds : l'agent se met à faire des ronds pour le temps restant de la trajectoire. . . . .	41
3.13	Autre exemple de comportement émergent provenant d'une autre série de l'expérience avec le comportement phototaxie sinusoïdale. L'agent tourne autour de la lampe sans s'en rapprocher, au contraire il s'en éloigne petit à petit. . . . .	42
3.14	Trajectoire de l'agent pendant le premier entraînement. Le point est la dernière position de l'agent. . . . .	43
3.15	Trajectoires de l'agent pendant le second entraînement. Le premier graphique correspond au motif <i>P1</i> , le second au motif <i>P2</i> . Le point est la dernière position de l'agent. . . . .	44
3.16	Nœuds et influences à la fin du premier entraînement. . . . .	45
3.17	Nœuds et dynamiques à la fin du second entraînement, pour les deux motifs, lorsque celui est dans le même espace sensorimoteur et à même valeur sensorielle (ici $s = 0$ ) que le premier entraînement. Les nœuds "hérités" du premier entraînement sont en bleus, ceux du second sont en vert. Dans les graphiques à droite, la trajectoire suivie par l'agent est le cercle vert. Nous pouvons voir que pour les deux motifs, seuls les nœuds du premier entraînement (bleus) sont présents à aux intersections des trajectoires, la densité de nœuds étant trop haute pour créer de nouveaux nœuds lors du passage de l'agent pendant le second entraînement. . . .	45
3.18	Résultat pour le motif <i>P1</i> lorsque les deux entraînements ont lieu dans le même espace sensorimoteur. Dans l'espace sensorimoteur, l'agent commence par aller vers la trajectoire du second entraînement mais finit toujours par suivre la dynamique du premier entraînement, plus rapidement lorsqu'il est relocalisé en (0.5, 0.5). Il suit cette dynamique jusqu'à la fin de l'expérience avec très peu d'écarts par rapport à la marche forcée. Son comportement stable est celui du premier entraînement (en bleu) : l'agent se dirige vers en bas à gauche en formant des boucles orientées en bas à droite. . . . .	47

3.19	Résultat pour le motif $P2$ lorsque les deux entraînements ont lieu dans le même espace sensorimoteur. Dans l'espace sensorimoteur, contrairement au motif $P1$ , le départ de l'agent est moins tranché, il n'est pas dirigé directement sur l'une ou l'autre des trajectoire d'entraînement mais oscille légèrement entre les deux. Une fois assez proche par contre il rejoint directement la dynamique résultant du premier entraînement et la suit jusqu'à la fin de l'expérience. Le comportement résultant est celui du premier entraînement : l'agent se dirige en bas à gauche (avec le motif $P2$ les boucles sont similaires pour les deux entraînements). . . . .	48
3.20	Nœuds et dynamique à la fin du second entraînement, pour les deux motifs, lorsque cet entraînement est dans un espace sensorimoteur différent de celui du premier entraînement mais toujours avec la même valeur sensorielle (ici $s = 0$ ). Nous pouvons voir que les nœuds couvrent bien les intersections contrairement à ceux de la figure 3.17, et que la dynamique est symétrique à celle du premier entraînement (voir figure 3.16). . . . .	49
3.21	Résultat pour le motif $P1$ lorsque les entraînements utilisent des espaces sensorimoteurs distincts. Dans l'espace sensorimoteur, l'agent alterne entre les trajectoires de chaque entraînement, soit en restant dans la partie interne lorsque la relocalisation est en $(0.5, 0.5)$ , soit au contraire en restant sur la partie externe avec la relocalisation en $0.8, 0.2$ ). Cela se traduit par un comportement différent de ceux des entraînements. Pour la première relocalisation l'agent effectue une petite ellipse qui est la combinaison des boucles. Pour la seconde relocalisation l'agent effectue une grande ellipse qui est la combinaison des arcs. . . . .	50
3.22	Résultat pour le motif $P2$ avec espaces sensorimoteurs distincts pendant les entraînements. Les deux relocalisations conduisent au même état stable en forme de croissant. Pour l'intersection du haut, les nœuds les plus récents et ayant donc un poids plus forts sont les nœuds du second entraînement ; pour celle du bas, c'est l'inverse, il s'agit de ceux du premier entraînement (voir les nœuds de la figure 3.16 et ceux de la figure 3.20b). Le comportement de l'agent est dans les deux cas un enchaînement d'arcs sans boucles et il conserve la direction du second entraînement.	51
3.23	Résultat pour le motif $P1$ lorsque les entraînements utilisent des espaces sensorimoteurs distincts et sont en plus dépendant de la valeur du senseur ( $s = 0$ pour le premier entraînement en bleu, $s = 1$ pour le second en vert). Les trajectoires sont similaires à celles de la figure 3.21, avec comme différence notable qu'il y a dans chaque cas une perturbation (petite boucle pour la relocalisation en $(0.5, 0.5)$ , écart avec la trajectoire bleue vers $m_2 = 0$ pour l'autre relocalisation). Dans les deux cas, cela est dû à l'activation du premier nouveau nœud qui est plus proche de l'agent que les nœuds d'entraînements. Avec l'activation successive des autres nœuds, l'agent retrouve un état stable proche de celui de la figure 3.21 . . . . .	53
3.24	Résultat pour le motif $P2$ lorsque les entraînements utilisent des espaces sensorimoteurs distincts et sont en plus dépendant de la valeur du senseur ( $s = 0$ pour le premier entraînement en bleu, $s = 1$ pour le second en vert). Comme pour le motif $P1$ , l'activation des nouveaux nœuds entraînement une perturbation brusque dans la trajectoire motrices et les états stables sont en partie modifiés par rapport à ceux de la figure 3.22. Il en va de même pour le comportement de l'agent qui dépend directement de ces trajectoires motrices. . . . .	54

---

3.25	Résultats qui illustre que forcer le senseur plus d'un côté que de l'autre permet de favoriser tel ou tel entraînement et par conséquent son comportement associé. Les perturbations dans les trajectoires motrices sont dues à l'activation des nouveaux nœuds. Ces derniers sont plus proches de l'agent et avec un poids plus fort que ceux des entraînements. . . . .	55
3.26	Évolution de la dynamique avec un IDSM de deux dimensions motrices et seulement 4 nœuds. Les nœuds représentés par les flèches grises ont toujours une poids $N_w = 0$ . Sur la figure 3.26a, la dynamique est équilibrée. Dans les cas extrêmes (figure 3.26b), le nœud rouge devient comme inexistant à gauche, contre une renforcement d'influence à peine perceptible à droite. Enfin, la figure 3.26c montre l'affaiblissement progressif du nœud rouge, ce qui se traduit par l'agrandissement de la zone d'équilibre depuis le centre vers la droite. . . . .	57
3.27	$\omega(N_w)$ en fonction de $N_w$ avec $k_\omega = 0.025$ à gauche et $k_\omega = 0.00025$ à droite. Les courbes sont identiques (y compris avec celle du second graphique de la figure 3.3), c'est l'ordre de grandeur de l'axe des abscisses qui change. . . . .	58
3.28	Affichage de la densité $\phi_{\mathcal{N}}$ pour un IDSM en une dimension, avec $k_\omega = 0.025$ à gauche et $k_\omega = 0.00025$ à droite. Le poids du nœud en 0.9 vaut $-2000$ (en bleu), $-100$ (orange), $0$ (vert), $100$ (rouge) et $2000$ (violet). Le poids du premier nœud en 0.25 reste à 0. Avec un faible $k_\omega$ , un nœud sera moins renforcé et moins affaibli pour une même variation de poids qu'avec un $k_\omega$ plus élevé. . . . .	59
3.29	Dynamiques en fin d'expérience avec $k_\omega = 0.25$ à gauche et $k_\omega = 0.00025$ à droite. Le choix de $k_\omega = 0.25$ est nécessaire pour voir un impact, les poids ne pouvant pas descendre en dessous de $-40$ ( $-1$ par unité de temps dans le pire des cas, voir l'équation 3.5). Le graphique de droite est identique à celui de la figure 3.5c. Sur le graphique de gauche, la trajectoire rouge s'éloigne légèrement de la trajectoire d'entraînement (en bleu). La densité devient assez faible entre chaque passage de l'agent pour qu'il recrée un nouveau nœud, renforçant le décalage. . . . .	59
3.30	Impact de $k_d$ sur la fonction $d$ et sur la densité $\phi_{\mathcal{N}}$ , avec $k_d = 30$ en bleu, $500$ en orange, $1000$ en vert, $2000$ en rouge et $10000$ en violet. Plus le $k_d$ augmente plus la zone d'influence du nœud rétrécit, et inversement. Sur le graphique de droite, les lignes rouges verticales pointillées indiquent la position des nœuds (dix nœuds placés aléatoirement avec des poids aléatoires dans $[-100, 0]$ ). La ligne verte horizontale pointillée indique $k_t$ le seuil de densité minimale pour la création de nouveaux nœuds. . . . .	60
3.31	Variation de $k_d$ sur le comportement cercle avec une durée de 60 unités de temps. $k_d = 30$ à gauche, $2000$ au centre et $10000$ à droite. Plus le $k_d$ est faible, plus la zone d'influence des nœuds est étendue. À gauche, le déficit de nœuds dans le quartier en bas à droite du cercle renforce l'influence du dernier nœud bleu en bas, ce qui déforme la trajectoire motrice et par conséquent le comportement de l'agent (cercle glissant). Au centre, avec $k_d = 2000$ , l'ensemble est similaire qu'avec $k_d = 1000$ (voir figure 3.5). Dans le dernier cas, avec $k_d = 10000$ , la zone d'influence des nœuds devient tellement faible qu'elle est considérée nulle au centre de l'espace sensorimoteur. Placer l'agent dans cette zone (ici en $(0.51, 0.5)$ ) bloque sa commande motrice actuelle (variation de $\mu$ nulle) et donc son comportement en cours : en l'occurrence, il se déplace légèrement vers la droite. . . . .	61

3.32	Exemples du comportement cercle sans délai avant activation des nœuds ( $\tau = 0$ ). L'écart d'influence entre le nœud nouvellement créé, nécessairement le plus proche, peut être suffisant pour changer le comportement résultant de l'agent. C'est le cas dans l'exemple avec la relocalisation en extérieur du cercle (0.13, 0.17). . . . .	62
3.33	Comparaison des nœuds créés entre à gauche avec une tentative de création de nœud dès que possible ( $\Delta t_{nodes} = \Delta t = 0.1$ ) et à droite qu'une fois par unité de temps ( $\Delta t_{nodes} = 1$ ). La faible couverture de l'entraînement dans le graphique de droite conduit à la fois à une influence plus forte des premiers nœuds créés pendant l'IDSM, et à la fois globalement à plus de nœuds créés sur le cercle. Ces derniers renforcent l'erreur au détriment de l'entraînement. . . . .	63
3.34	Comparaison entre à gauche avec une tentative de création de nœud dès que possible ( $\Delta t_{nodes} = \Delta t = 0.1$ ) et à droite qu'une fois par unité de temps ( $\Delta t_{nodes} = 1$ ). La trajectoire motrice est moins stable, surtout dans le quart en bas à gauche, ce qui entraîne un comportement de cercle glissant. . . . .	64
3.35	Impact du paramètre $k_t$ , avec $k_t = 0.01$ à gauche et $k_t = 10$ à droite. Un faible $k_t$ veut dire une saturation plus rapide, donc peu de création de nœuds et un système plus rigide. Un fort $k_t$ à l'inverse rend la saturation en nœuds plus lente à atteindre et un système plus plastique. Sur le second graphique de la figure 3.35b, le cercle rouge s'épaissit car l'influence des nœuds du centre est prise en compte lors de la création des nouveaux nœuds du cercle, poussant la trajectoire motrice plus loin. . . . .	65
3.36	Dynamiques en fin d'expérience, à gauche en limitant la commande motrice $\mu \in [0, 1]$ , à droite avec $\mu \in [-\infty; +\infty]$ . La valeur de $\mu$ est tracée en mauve, elle n'est pas visible à gauche car strictement égale à la position sensorimotrice de l'agent (en rouge). Sur le graphique de gauche, l'agent "récupère" la trajectoire d'entraînement dès que possible (la dynamique étant équilibrée, dès $(m_1, m_2) = (0.5, 0)$ ). À droite, il ne superpose la trajectoire d'entraînement qu'après $(m_1, m_2) = (1, 0.5)$ et un décalage permanent reste visible entre trajectoires en bas à droite du cercle à cause de l'influence des nouveaux nœuds. . . . .	66
4.1	Illustration des comportements attendus avec ou sans habitude. Hors stimulation, le comportement prédéfini de l'agent est de monter. Lorsque le stimulus est perçu, son comportement prédéfini est de descendre. Avec l'habitude, nous nous attendons à ce que l'agent apprenne progressivement à se comporter comme hors stimulation lorsque la stimulation est continue ou répétée (c'est-à-dire monter à la place de descendre). . . . .	70
4.2	Le facteur d'apprentissage $f_a$ comme fonction de la différence entre la valeur sensorielle brute $s_b(\mathbf{x}_E)$ et la mémoire sensorielle $\delta$ . . . . .	72
4.3	IDSM avec deux nœuds utilisé dans les expériences sur l'habitude. Les flèches colorées représentent les nœuds, les flèches noires indiquent la dynamique motrice : la composante motrice $[\mathbf{x}_{SM}]^\mu$ tend vers 0 lorsque le stimulus est perçu ou vers 1 lorsque le stimulus n'est pas perçu. Les comportements liés au moteur sont que l'agent descend plus ou moins vite lorsque $[\mathbf{x}_{SM}]^\mu$ est dans $[0; 0.5[$ , l'agent est à l'arrêt lorsque $[\mathbf{x}_{SM}]^\mu$ vaut 0.5 et l'agent monte plus ou moins vite lorsque $[\mathbf{x}_{SM}]^\mu$ est dans $]0.5, 1]$ . . . . .	75

---

4.4	Mécanisme de mémoire sensorielle pour les quatre schémas de stimulation différents pour tester s'il y a habituation et récupération spontanée. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice $\mu$ , la mémoire sensorielle de l'agent $\delta$ . Dans chacun des quatre cas, l'agent apprenant s'habitue rapidement et cesse d'exprimer la réponse motrice (descente) au stimulus pour progressivement se déplacer vers le haut. Le schéma de stimulation a un impact sur la vitesse et l'amplitude de l'habituation. . . . .	76
4.5	Mécanisme de pénalité sur l'énergie des nœuds pour les quatre schémas de stimulation pour tester s'il y a habituation et récupération spontanée. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la composante motrice $[\mathbf{x}_{SM}]^\mu$ et la commande motrice $\mu$ , les valeurs de surprise, l'énergie et le $\xi_N$ des nœuds. Dans le premier graphique (a), le manque d'évènements surprenants ne permet pas à l'agent de s'habituer, contrairement au dernier graphique (d) où l'habituation a lieu. Dans les graphiques (b,c), nous observons une habituation inverse (le timing des surprises entraîne une diminution de l'énergie de $N^2$ ). . . . .	77
4.6	Mécanisme de pénalité sur le poids des nœuds pour les quatre schémas de stimulation pour tester s'il y a habituation et récupération spontanée. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la composante motrice $[\mathbf{x}_{SM}]^\mu$ et la commande motrice $\mu$ , les valeurs de surprise, les poids et le $\xi_N$ des nœuds. Dans le premier graphique (a), le manque d'évènements surprenants ne permet pas à l'agent de s'habituer. Du graphique (b) à (d), nous observons une transition progressive de l'habituation inverse à l'habituation lorsque la pénalité passe de $N^2$ à $N^1$ . . . . .	78
4.7	Trajectoires sensorimotrices montrant l'impact du rapport entre l'activation et l'inactivation du stimulus. Les flèches épaisses du haut et du bas représentent les deux nœuds. Les flèches horizontales fines représentent la trajectoire de l'agent dans l'espace sensorimoteur, elles devraient toujours être soit à $s = 1$ lorsque le stimulus est activé, soit à $s = 0$ lorsque le stimulus est désactivé, mais sont espacées pour une meilleure lisibilité. Les flèches verticales courbes montrent le saut dans la trajectoire dû à l'activation ou à l'inactivation du stimulus. Les flèches avec de grands tirets représentent le cycle dans lequel l'agent termine. L'agent commence toujours au point noir (également espacé verticalement pour plus de lisibilité). Les graphiques du haut représentent les trajectoires de l'agent témoin, ceux du bas sont les trajectoires de l'agent avec un modèle de pénalité des nœuds. Le graphique en bas à droite illustre l'habituation, tandis que le graphique en bas à gauche illustre l'habituation <i>inverse</i> . . . . .	81
4.8	Test de l'effet de l'intervalle interstimulus avec le mécanisme de mémoire sensorielle. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice $\mu$ , la mémoire sensorielle de l'agent $\delta$ . Plus l'intervalle interstimulus augmente (en gardant un temps d'activation constant), plus l'habituation diminue. Notons que la durée des expériences varie de 40 à 70 unités de temps. . . . .	82

4.9	<p>Test de l'effet de l'intervalle interstimulus avec le mécanisme de pénalité sur l'énergie des nœuds. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice <math>\mu</math>, les valeurs de surprise, l'énergie et le <math>\xi_N</math> des nœuds. Pour le schéma (a), nous observons une habituation inverse. Pour les tracés (b) à (d), lorsque l'intervalle interstimulus augmente, l'habituation est moins rapide à se mettre en place. Notons que la durée des expériences varie de 40 à 70 unités de temps. . . . .</p>	83
4.10	<p>Test de l'effet de l'intervalle interstimulus avec le mécanisme de pénalité sur le poids des nœuds. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice <math>\mu</math>, les valeurs de surprise, les poids et le <math>\xi_N</math> des nœuds. Pour tous les schémas de stimulation sauf (a), l'habituation a lieu mais repose sur une interaction plus subtile entre les pénalités <math>N^1</math> et <math>N^2</math> que les autres expériences. Voir le texte principal pour plus de détails. Notons que la durée des expériences varie de 40 à 70 unités de temps. . . . .</p>	84
4.11	<p>Combinaison des modèles avec mémoire sensorielle et avec pénalité sur l'énergie des nœuds pour tester s'il y a habituation dans chacun des schémas de stimulation. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice <math>\mu</math>, la mémoire sensorielle de l'agent <math>\delta</math> et la valeur des surprises, l'énergie et le <math>\xi_N</math> des nœuds. Pour chaque schéma, l'habituation est soit identique soit plus forte que celle observée avec chaque modèle individuellement. Le cas particulier de l'habituation inverse avec la pénalité sur l'énergie des nœuds (voir figure 4.5) devient une habituation "ordinaire". . . . .</p>	87
4.12	<p>Combinaison des modèles avec mémoire sensorielle et avec pénalité sur le poids des nœuds pour tester s'il y a habituation dans chacun des schémas de stimulation. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice <math>\mu</math>, la mémoire sensorielle de l'agent <math>\delta</math> et la valeur des surprises, le poids et le <math>\xi_N</math> des nœuds. Pour chaque schéma, l'habituation est soit identique soit plus forte que celle observée avec chaque modèle individuellement. Le cas particulier de l'habituation inverse avec la pénalité sur l'énergie des nœuds (voir figure 4.6) devient une habituation "ordinaire". . . . .</p>	88
4.13	<p>Illustration de la chaîne d'événements entre la mémoire sensorielle <math>\delta</math>, la valeur du senseur habitué et l'amplitude de la surprise avec a) le zoom sur la figure 4.11c et b) le zoom sur la figure 4.5c comme référence. Sur a), la surprise est plus forte au début du stimulus et plus faible à la fin du stimulus en raison de l'accumulation de la mémoire sensorielle <math>\delta</math>. . . . .</p>	89
4.14	<p>Impact des modèles avec pénalité des nœuds sur la dynamique de l'espace sensorimoteur, en modifiant soit l'énergie du nœud <math>N^1</math> <math>N_e</math> (a,b) soit le poids du nœud <math>N^1</math> <math>N_w</math> (c,d). L'influence de <math>N^1</math> diminue à la fois avec <math>N_e</math> et <math>N_w</math>, mais dans le cas du poids l'influence de <math>N^2</math> augmente de manière très significative. . . . .</p>	92

---

4.15	Agent témoin avec le schéma (b) utilisant l'IDSMS original. De haut en bas : la trajectoire de l'agent témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice $\mu$ . Avec la dynamique modifiée par les nouveaux nœuds, sans borner $\mu$ , l'agent se comporte comme si le stimulus était actif de manière permanente lors phases de stimulation une fois les nœuds actifs. . . . .	95
4.16	Modèle avec la mémoire sensorielle basé sur l'IDSMS original avec le schéma (a) pour tester s'il y a habituation. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice $\mu$ , la mémoire sensorielle de l'agent $\delta$ . L'habituation et la récupération spontanée sont toujours observées même si de nouveaux nœuds sont créés et les poids des nœuds continuellement mis à jour. . . . .	96
4.17	Modèle avec pénalisation de l'énergie utilisant l'IDSMS original avec le schéma (d) pour tester s'il y a habituation. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice $\mu$ , la mémoire sensorielle de l'agent $\delta$ et la valeur des surprises, l'énergie et le $\xi_N$ des nœuds. L'habituation et la récupération spontanée sont toujours observées avec la création de nouveaux nœuds et la modification continue des poids. . . . .	97
4.18	Modèle avec mémoire sensorielle et pénalisation de l'énergie utilisant l'IDSMS original avec le schéma (d) pour tester s'il y a habituation. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice $\mu$ , la mémoire sensorielle de l'agent $\delta$ et la valeur des surprises, l'énergie et le $\xi_N$ des nœuds. L'habituation et la récupération spontanée sont toujours observées avec la création de nouveaux nœuds, la modification continue des poids et la combinaison de deux modèles. . . . .	97
4.19	Illustration des comportements attendus avec ou sans sensibilisation spécifique au site. Hors stimulation, le comportement prédéfini de l'agent est de monter. Lorsque le stimulus est perçu, son comportement prédéfini est de descendre. Avec la sensibilisation, nous nous attendons à ce que l'agent apprenne progressivement à répondre de plus en plus fortement au stimulus lorsque la stimulation est continue (a) ou répétée (b), c'est-à-dire de descendre plus vite que s'il était incapable de sensibilisation. . . . .	99
4.20	IDSMS avec deux nœuds utilisé dans les expériences sur la sensibilisation spécifique au site. Les flèches colorées représentent les nœuds, les flèches noires indiquent la dynamique motrice : la composante motrice $[\mathbf{x}_{SM}]^\mu$ tend vers 0 lorsque le stimulus est perçu ou vers 1 lorsque le stimulus n'est pas perçu. Les comportements liés au moteur sont que l'agent descend plus ou moins vite lorsque la composante motrice $[\mathbf{x}_{SM}]^\mu$ est dans $[0; 0.5[$ , l'agent est à l'arrêt lorsque $[\mathbf{x}_{SM}]^\mu$ vaut 0.5 et l'agent monte plus ou moins vite lorsque $[\mathbf{x}_{SM}]^\mu$ est dans $]0.5, 1]$ . . . . .	102
4.21	Mécanisme de mémoire sensorielle pour les quatre schémas de stimulation pour tester s'il y a sensibilisation spécifique au site. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice $\mu$ , la mémoire sensorielle de l'agent $\delta$ . Comme pour l'habituation, le modèle avec mémoire sensorielle permet une meilleure sensibilisation spécifique au site lorsque la stimulation est longue plutôt que brève. . . . .	103

4.22 Mécanisme de bonus sur l'énergie des nœuds pour les quatre schémas de stimulation pour tester s'il y a sensibilisation spécifique au site. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la composante motrice  $[\mathbf{x}_{SM}]^\mu$  et la commande motrice  $\boldsymbol{\mu}$ , les valeurs de surprise, l'énergie et le  $\xi_N$  des nœuds. Dans le premier graphique (a), le manque d'évènements surprenants ne permet pas à l'agent de se sensibiliser. Comme pour l'habituation inverse, les schémas (b) et (c) montrent une sensibilisation inverse (renfort du comportement non stimulé plutôt que du comportement réponse à la stimulation). Le schéma (d) montre une sensibilisation avant que l'influence des deux nœuds redeviennent équilibrés. . . . . 105

4.23 Mécanisme de bonus sur le poids des nœuds pour les quatre schémas de stimulation pour tester s'il y a sensibilisation spécifique au site. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la composante motrice  $[\mathbf{x}_{SM}]^\mu$  et la commande motrice  $\boldsymbol{\mu}$ , les valeurs de surprise, l'énergie et le  $\xi_N$  des nœuds. Dans le premier graphique (a), le manque d'évènements surprenants ne permet pas à l'agent de se sensibiliser. Comme pour l'habituation inverse, les schémas (b) et (c) montrent une sensibilisation inverse (renfort du comportement non stimulé plutôt que du comportement réponse à la stimulation). Le schéma (d) montre une sensibilisation à chaque phase de stimulation. . . . . 106

4.24 Combinaison des modèles avec mémoire sensorielle et avec pénalité sur l'énergie des nœuds pour tester s'il y a sensibilisation spécifique au site dans chacun des schémas de stimulation. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\boldsymbol{\mu}$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. Contrairement à l'habituation où la combinaison transformait l'habituation inverse en habituation "régulière", ici la combinaison permet au modèle avec énergie de renforcer les deux nœuds plus tôt et plus fortement, rendant plus rapidement l'agent très sensible aux variations de stimulation (activation et désactivation), en particulier avec le schéma (b). . . . . 108

4.25 Combinaison des modèles avec mémoire sensorielle et avec pénalité sur le poids des nœuds avec le schéma (b). De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\boldsymbol{\mu}$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, le poids et le  $\xi_N$  des nœuds. Il y a sensibilisation inverse, la mémoire sensorielle permet d'avoir un impact plus important sur le poids des nœuds et ainsi de rendre le comportement de l'agent plus marqué qu'avec le modèle avec bonus sur les poids seul. . . . . 109

4.26 Modèle avec la mémoire sensorielle basé sur l'IDSMS original lors du schéma (a) pour tester s'il y a sensibilisation spécifique au site. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\boldsymbol{\mu}$ , la mémoire sensorielle de l'agent  $\delta$ . La sensibilisation spécifique au site est toujours observable même si de nouveaux nœuds sont créés et les poids des nœuds continuellement mis à jour. . . 110



- 
- 4.27 Modèle avec bonus sur l'énergie basé sur l'IDSM original lors du schéma (c) et (d) pour tester s'il y a sensibilisation spécifique au site. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. Avec le schéma (d), la sensibilisation est toujours observée avec la création de nouveaux nœuds et la modification continue des poids, ainsi que la sensibilité de l'agent aux modifications du stimulus (activation et désactivation). Nous pouvons aussi voir avec le schéma (c) qu'avec la création de nœuds, la trajectoire motrice évolue dans une dynamique très instable de la fin de la phase 1 à la phase 2, mais pas du tout en phase 4. . . 111
- 4.28 Modèles avec mémoire sensorielle, bonus sur l'énergie et combinaison des deux lors du schéma (b) et utilisant l'IDSM original pour tester s'il y a sensibilisation spécifique au site. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. Globalement, les résultats sont similaires à ceux avec l'IDSM simplifié : sensibilisation pour le modèle avec mémoire sensorielle, sensibilisation inverse et sensibilité à l'activation et désactivation du stimulus pour le modèle avec énergie, et renfort de ce dernier phénomène lorsque les modèles sont combinés. . . . . 113
- 4.29 Illustration des comportements attendus avec ou sans pseudo-conditionnement. Lorsque le stimulus inconditionnel  $SI$  ( $US$  sur la figure) est perçu, le comportement prédéfini de l'agent est de descendre. Hors stimulation, son comportement prédéfini est de monter. Lorsque le stimulus conditionnel  $SC$  ( $CS$  sur la figure) est perçu, ce stimulus étant neutre, il se comporte comme hors stimulation et monte. Avec le pseudo-conditionnement, nous nous attendons à ce qu'après une exposition à stimulus inconditionnel  $SI$ , l'agent apprenne à répondre de la même manière lorsqu'il perçoit le stimulus conditionnel  $SC$ , c'est-à-dire qu'il descende à la place de monter, puis qu'il se remette à monter après un certain temps sans exposition au stimulus inconditionnel  $SI$ . Les stimulations peuvent être continues (a) ou répétées (b). . . . . 115
- 4.30 Variation de la commande motrice ( $d\mu$ ) dans le sous-espace sensorimoteur  $m \times s_1$  d'un IDSM avec deux nœuds utilisé dans les expériences sur le pseudo-conditionnement. Les grandes flèches sont les nœuds, la base étant le  $N_p$  et la longueur de la flèche représentant le  $N_v$ . a) : le masque de  $N^1$  est  $N_m = [1, 1, 1]$ . b) : avec un masque  $N_m = [0.5, 0.5, 0.5]$ , cela conduit au renforcement de son influence face à  $N^2$ . . . 117
- 4.31 Résultats pour les schémas (a) et (b) pour tester s'il y a pseudo-conditionnement. De haut en bas : trajectoires des agents apprenant et témoin, valeurs des senseurs, valeurs de la composante motrice  $[x_{SM}]^\mu$  et valeurs des masques. Lors de la troisième phase (après 40u.t.), la généralisation plus forte de  $N^1$  par rapport à  $N^2$  (voir le graphique des masques) conduit à un comportement pseudo-conditionné de l'agent : il apprend à se déplacer vers le bas lorsque le  $SC$  est perçu. . . . . 118

4.32 Résultats pour les schémas (c) et (d) pour tester s'il y a pseudo-conditionnement. De haut en bas : trajectoires des agents apprenant et témoin, valeurs des senseurs, valeurs de la composante motrice  $[x_{SM}]^u$  (attention à l'échelle par rapport aux schémas (a) et (b), ici la valeur minimale est 0.5) et valeurs des masques. Ces schémas mettent en évidence que pendant la phase 2 (entre 20u.t. et 40u.t.), la généralisation des masques conduit à des contributions similaires des deux nœuds et donc à l'immobilité de l'agent. En phase 2 et 3, les valeurs des masques inhibent fortement l'expression du comportement de descente, le moteur de l'agent met seulement plus ou moins de temps à atteindre son maximum. . . . . 119

# Introduction

Cette thèse se situe dans les domaines de l'intelligence artificielle et des Sciences Cognitives sur la thématique de l'autonomie pour un agent artificiel.

La notion d'autonomie est une notion très large et dont il est difficile de trouver une définition formelle universelle. En particulier, nous pouvons remarquer que l'autonomie d'un animal est une notion relative qui dépend notamment du contexte dans lequel cet animal est placé. Par exemple, si nous prenons un humain et un dauphin, ce sont tous deux des animaux que l'on peut considérer comme autonomes dans leur environnement naturel. Par contre, si nous les plaçons l'un dans l'environnement de l'autre, disons le dauphin au milieu de Paris et l'humain au milieu de l'Atlantique, cette autonomie peut être remise en question. L'autonomie d'un certain organisme pourrait alors être décrite comme une caractéristique résultante d'une interaction entre, au moins, l'environnement dans lequel évolue l'organisme, les changements possibles dans cet environnement auxquels l'organisme doit être adapté ou s'adapter spontanément, le mode de vie de l'organisme (quelles tâches il doit accomplir pour se maintenir en vie par exemple), la morphologie de l'organisme et les apprentissages dont il dispose. Cette liste est bien entendu discutable, ne serait-ce que sur les interdépendances entre ses différents éléments, mais cela permet d'illustrer naïvement et rapidement la complexité derrière une notion telle que celle d'autonomie.

Ainsi, si l'on conçoit intuitivement l'autonomie comme le fait de pouvoir évoluer et survivre à partir de ses propres actions ou décisions, le concept est délicat à rendre opératoire, à décrire avec des équations ou à programmer au sein d'un agent artificiel. S'agit-il d'un problème d'optimisation, auquel cas quelle serait la fonction à optimiser ? S'agit-t-il de maintenir des variables dans certains intervalles pour assurer la bonne marche de l'individu ? Peut-on diviser ou structurer cette propriété globale en plusieurs caractéristiques moins complexes et plus faciles à décrire ?

Une approche possible pour partir d'une définition plus "concrète" de la notion d'autonomie est de s'intéresser au sens premier du mot. Si nous cherchons l'étymologie du terme *autonome*, *autónomos* en grec ancien, la signification du mot est "qui est capable de faire ses propres lois" (*autós* + *némô*). Nous pouvons alors étendre cette définition simple à un agent artificiel en considérant qu'il gagne en autonomie s'il est capable de modifier de "lui-même" les comportements sensorimoteurs "réactifs" ou "réflexes" (réponse motrice à une situation sensorielle perçue) qu'il possède initialement et qui font office de comportements innés ou, plus largement, de compor-

tements précédemment acquis. Il restera bien sûr à mieux préciser ce que nous entendons par "comportement", ce que nous ferons plus avant dans cette thèse.

Nos travaux se concentrent sur ces modifications de comportements que l'on peut décliner sous trois formes : la modulation de comportements face à certaines situations de l'environnement, l'exploration autour des comportements déjà acquis et l'acquisition (mémorisation) de nouveaux comportements. Nous entendons par modulation de comportements que l'agent soit capable d'adapter ses comportements au contexte, selon ce qu'il perçoit ou ce qu'il fait. Par exemple, il devrait pouvoir traiter le cas où un obstacle anodin se trouve sur son chemin, soit en l'ignorant (modulation de la réponse à la perception de l'obstacle), soit en le contournant (modulation de l'exploration en cours). L'agent doit pouvoir "explorer" autour des comportements prédéfinis ou acquis. Nous entendons par là qu'il ne reproduise pas toujours exactement le même mouvement lorsqu'il se comporte (à travers un biais aléatoire par exemple), lui permettant alors d'expérimenter à la marge de son activité motrice déjà connue. Enfin, l'agent doit être en mesure d'acquérir de nouveaux comportements. Si les comportements résultants de son adaptation au contexte ou à son exploration motrice semblent intéressants, il doit être capable de les mémoriser en tant que comportement acquis exprimables au même titre qu'un comportement prédéfini.

Nous nous intéressons alors à la notion d'apprentissage, qui couvre en partie nos besoins concernant les modifications de comportements, en particulier concernant l'adaptation et l'acquisition telles que nous les avons évoquées.

Il existe déjà de nombreux travaux autour de l'apprentissage en intelligence artificielle, qui convoquent des concepts comme par exemple l'apprentissage supervisé, l'apprentissage non-supervisé ou l'apprentissage par renforcement. L'apprentissage supervisé cherche à estimer une fonction à partir d'un ensemble de données *étiquetées* au préalable, là où l'apprentissage non supervisé utilise des données brutes, sans informations supplémentaires, pour les catégoriser ou les regrouper par similarité. Dans l'apprentissage par renforcement, l'agent apprend de manière itérative, par essais et erreurs, à travers des récompenses associées aux actions qu'il peut effectuer.

Cependant, au sens de la biologie, ces apprentissages apparaissent déjà assez complexes, assez "évolués". Le conditionnement opérant, source d'inspiration de l'apprentissage par renforcement, ou la plasticité neuronale, qui sous-tend tous les travaux sur les réseaux de neurones formels, sont des formes d'apprentissages qui n'ont été observés que dans un nombre relativement restreint d'espèces animales.

À l'inverse, les travaux en intelligence artificielle ne se sont que peu penchés sur les mécanismes d'apprentissage non-associatifs. Pourtant, en biologie, ces apprentissages sont considérés comme les plus simples et sont largement présents dans le vivant, y compris chez des organismes dénués de système nerveux. Cette forte représentation plaide en faveur d'un intérêt évolutif de ces apprentissages, qui pourraient être des mécanismes primordiaux d'adaptations face aux changements ponctuels dans l'environnement de l'organisme. Ces apprentissages pourraient aussi être les fondements des apprentissages de plus haut niveau, en particulier de l'apprentissage associatif qu'est le conditionnement classique (aussi appelé conditionnement Pavlovien).

L'hypothèse que les apprentissages non-associatifs pourraient être nécessaires à une autonomie face à de nombreux types de milieu ou de modes de vie nous intéresse particulièrement. Nous précisons alors la problématique de cette thèse ainsi : comment s'inspirer des apprentissages non-associatifs tels que décrits et compris par la biologie pour élaborer des mécanismes opérationnels

---

pour des agents artificiels? L'objectif serait alors d'explorer et d'analyser les capacités offertes pas ces nouveaux modèles.

Dans la littérature, les travaux en intelligence artificielle ou en robotique existants qui utilisent les apprentissages non-associatifs sont essentiellement basés sur des architectures neuronales. De telles architectures ont l'inconvénient d'être souvent difficiles à analyser car elles mettent en jeu des structures complexes où de nombreux mécanismes sont associés et interdépendants.

Cependant, il existe de nombreuses autres architectures d'agents artificiels. L'une d'elle, l'*Iterant Deformable Sensorimotor Medium* (IDSM), a retenu l'attention au sein de l'équipe Biscuit au cours de travaux précédents. Cet IDSM repose sur un espace sensorimoteur de l'agent, où les comportements sont représentés en tant que trajectoires sensorimotrices et où ils sont mémorisés partiellement à l'aide de structures peuplant cet espace appelées *nœuds*. Ce niveau de description, appelé mésoscopique, est intermédiaire entre un niveau d'activité neuronale et un niveau d'actions macroscopiques. Un avantage de se placer à un niveau plus abstrait que celui des architectures neuronales est de permettre des analyses plus simples ou plus intuitives. De plus, la représentation des comportements dans l'espace sensorimoteur permet une représentation continue des actions réalisables par l'agent par rapport à un ensemble prédéfini d'actions discrétisées. Enfin, ce modèle a été conçu spécifiquement pour réfléchir sur la notion d'habitude en tant que tendance à ré-exprimer une trajectoire sensorimotrice déjà rencontrée et les questions sur la modulation et l'exploration autour de comportements existants n'y ont pas été explorées.

Nous avons donc décidé de nous appuyer sur cet IDSM pour nos travaux sur un agent artificiel doté des apprentissages non-associatifs. Nous concrétisons notre problématique autour de la question suivante : en basant notre inspiration sur les différents apprentissages non-associatifs, quels mécanismes pourraient être ajoutés au cadre de l'IDSM afin de permettre à un agent artificiel les modifications de comportements que sont l'adaptation du comportement au contexte environnemental, l'exploration autour d'un comportement déjà acquis et l'acquisition de nouveaux comportements ?

Pour apporter une réponse à cette problématique, ce manuscrit se structure ainsi. Dans le chapitre 2, nous revenons d'abord sur les notions de *comportement* et d'*apprentissage* pour voir sur quelles définitions formelles de ces concepts nous pourrions nous reposer. Nous dressons ensuite un état de l'art sur différents apprentissages observés en biologie, en particulier sur les apprentissages non-associatifs, puis sur les travaux connexes en modélisation et intelligence artificielle. Cela permet de dégager les propriétés de trois apprentissages qui nous ont particulièrement intéressés : l'habituation, la sensibilisation spécifique au site et le pseudo-conditionnement. Dans le chapitre 3, nous présentons formellement le modèle IDSM. Nous illustrons son fonctionnement et ses caractéristiques à travers des expériences reproduites de l'article de M.Egbert (M. D. EGBERT et BARANDIARAN, 2014) ainsi qu'une première contribution de notre part sous la forme d'expériences de combinaisons de comportements basées sur ce modèle. Nous dressons alors un premier bilan sur l'IDSM et ses propriétés. Le chapitre 4 présente nos principales contributions. Il s'agit des différents modèles que nous proposons qui sont couplés à un IDSM simplifié pour doter un agent artificiel d'habituation, de sensibilisation spécifique au site ou de pseudo-conditionnement selon le modèle. Ces modèles sont présentés puis testés et analysés par le biais de simulations dans des environnements virtuels contrôlés et simples. Nous concluons ce chapitre par un bilan sur nos modèles. Enfin, le chapitre 5 dresse un bilan global de la thèse et explore différentes perspectives.



## 2

# État de l'art

## 2.1 Biologie

Nos travaux de thèse portent sur la construction de modèles informatiques permettant l'acquisition de comportement. La biologie peut, de ce point là, être une source d'inspiration. En particulier, dans ce chapitre, après être revenu plus formellement sur la notion de comportement, nous présentons différentes formes d'apprentissage observées dans le monde du vivant et mettons en avant les apprentissages non-associatifs qui paraissent être universelles au sens où ils sont présents dans un grand nombre d'espèces. Ce chapitre se termine par une présentation des travaux antérieurs en modélisation informatique d'agents artificiels apprenant.

### 2.1.1 Comportements

La notion de comportement fait partie de celles que l'on utilise régulièrement dans le langage courant, que tout le monde comprend *a priori*, mais qui reste difficile à définir de manière générale.

Le dictionnaire Merriam-Webster donne une définition assez large de *behavior* :

« *Anything that an organism does involving action and response to stimulation.* »

À partir de là, nous pouvons voir rapidement avec un exemple le genre d'ambiguïté qui se pose avec le terme. En effet, imaginons qu'un animal trouve une baie correspondant à son régime alimentaire. Nous pouvons alors décrire la suite de comportements : il prend la baie, il la met dans sa bouche, il mâche, etc. Nous pouvons tout autant décrire le comportement : il mange. Et de manière encore plus gros grain : après s'être réveillé, il cherche de quoi se nourrir dans les alentours de sa tanière.

Dans les travaux présentés dans cette thèse, sur le terme général, nous resterons simplement sur l'idée d'interaction de l'agent avec son environnement, tout en nous permettant de l'employer à la fois pour une seule réponse (par exemple "après détection de la lumière il recule") et pour ce qu'il fait de manière globale (par exemple "s'il perçoit la lumière par intermittence à de multiples reprises, il recule de moins en moins").

Ici cependant, nous pouvons remarquer que l'observateur apporte un jugement subjectif. Il discrétise ce que fait l'animal (ou dans notre cas l'agent) en étapes qu'il perçoit et qu'il comprend.

De plus, le niveau de description de ces étapes est défini selon le contexte de l'observation. La notion de comportement est arbitraire, elle n'a pas de réalité physique pour l'organisme qui interagit dans un flux continu constant d'informations sensorielles. Si nous prenons l'exemple d'une chauve-souris en plein vol, nous pouvons décrire qu'elle vole car nous le voyons, mais sans connaissances extérieures nous ne pouvons savoir qu'elle émet des ultrasons car nous ne les percevons pas. De même, la granularité du comportement "vol", "bat des ailes" ou "contracte tel ou tel muscle" dépend de l'étude en court.

La formalisation de telles discrétisations est nécessaire pour l'étude scientifique du comportement animal ou plus généralement pour l'étude scientifique du comportement de n'importe quel organisme vivant.

Le domaine de la biologie qui s'occupe de manière globale du comportement est l'*éthologie*. C'est une science relativement récente, reconnue seulement depuis 1973, qui peut être brièvement définie comme "l'étude biologique du comportement". BURKHARDT, 1999 et MORENO et MUÑOZ-DELGADO, 2007 permettent d'avoir une idée de l'émergence de cette science depuis les travaux d'Aristote à nos jours. Nous en présentons ici les grandes lignes.

Le terme éthologie a été utilisé par Isidore Geoffroy-Saint-Hilaire en 1859 pour parler de l'étude des animaux vivants dans leur environnement naturel plutôt que morts en laboratoire (MORENO et MUÑOZ-DELGADO, 2007). Les bases formelles de l'éthologie ont été établies par Nikolaas Tinbergen et Konrad Lorenz à partir des années 1940. En particulier, Tinbergen orientera les éthologues à ancrer leurs travaux selon quatre axes : recherche de la cause mécanique du comportement, recherche de son intérêt fonctionnel ou adaptatif pour l'individu, recherche de la manière selon laquelle l'individu en est arrivé là (ontogénie, développement physiologique, apprentissage), recherche de la manière selon laquelle l'espèce en est arrivée là (phylogénie).

Cependant, il faudra attendre 1973 pour que l'éthologie soit reconnue en tant que discipline scientifique, avec l'attribution du prix Nobel de médecine et de physiologie à trois de ses représentants remarquables : Konrad Lorenz, Niko Tinbergen et Karl von Frisch.

Une science plus ancienne, la psychologie animale, était elle apparue dès la fin du XIX<sup>e</sup> siècle pour traiter du comportement, de l'apprentissage et de l'intelligence chez les animaux.

Au cours du XX<sup>e</sup> siècle, l'éthologie et la psychologie animale se sont essentiellement développées de manière indépendante (RICHARDS, 1987 ; SHETTLEWORTH, 2001). Shettleworth donne plusieurs explications sur le manque de communication entre les deux domaines. En particulier, les éthologues se sont focalisés sur l'étude des comportements innés sur le terrain, à travers le prisme de la théorie de l'évolution. Les psychologues animalier, quant à eux, se préoccupaient peu de la théorie de l'évolution, voire certains étaient contre, et travaillaient surtout en laboratoire sur les comportements appris (ou acquis) d'un nombre restreint d'espèces. À partir des années 80, il y a eu de plus en plus d'échanges entre l'éthologie et la psychologie animale et de sous-domaines les reliant directement comme l'éthologie cognitive ou la psychologie évolutive par exemple.

Ce que nous retiendrons comme définition générale pour comportement est une réponse de l'organisme (ou de l'agent dans le cas concerné par cette thèse) à l'environnement. L'environnement ici est l'ensemble des stimuli présents ou passés perceptibles par l'organisme (ou l'agent). Dans la suite, pour nos agents, nous utiliserons la définition de comportement inné comme un comportement non-appris, "pré-câblés", à la différence du comportement acquis qui lui est appris.



Nous allons ensuite voir plus en détail le concept d'apprentissage, et les définitions de différents apprentissages rencontrés dans le vivant.

### 2.1.2 Apprentissages

Comme pour le comportement, l'apprentissage est une notion utilisée très régulièrement sans pour autant avoir une définition universelle qui fasse l'unanimité. L'article de Barron et co-auteurs (BARRON et al., 2015) regroupe de nombreuses définitions en provenance de plusieurs domaines scientifiques (psychologie, psychologie cognitive, neuroscience, écologie comportementale, apprentissage automatique), classées en trois catégories : traitement de l'information ou de l'expérience (définitions conceptuelles), changement comportemental, changement dans les mécanismes du comportement. Barron et coauteurs considèrent que la plupart de ces définitions entrent dans un cadre théorique commun : l'apprentissage comme une mise à jour structurée des propriétés du système basée sur le traitement de nouvelles informations.

Dans un article relativement récent, De Houwer, Barnes-Holmes et Moors (DE HOUWER, BARNES-HOLMES et MOORS, 2013) discutent des avantages et inconvénients de différentes définitions historiques avant de proposer leur propre définition inspirée des travaux de Skinner (B. SKINNER, 1938 ; Burrhus Frederic SKINNER, 1984). Les définitions historiques de l'apprentissage s'opposent entre deux catégories : les définitions *fonctionnelles* et les définitions *mécanistes*.

La définition *fonctionnelle* de l'apprentissage la plus répandue est le changement de comportement dû à l'expérience. Cette définition permet déjà d'exclure tout changement de comportement faisant suite à autre chose que l'expérience, comme par exemple un changement dû à la génétique, à la fatigue, ou encore à la motivation. Cependant, cette définition ne prend pas en compte des cas où l'on peut considérer qu'il y a apprentissage, alors que cela ne se traduit pas par une modification du comportement. Ce cas est illustré, par exemple, par les travaux de Tolman et Honzik (TOLMAN et HONZIK, 1930) dans lesquels ils mettent en avant un apprentissage latent avec des rats explorant un labyrinthe. Les rats qui ont eu l'occasion d'explorer préalablement le labyrinthe même sans nourriture sont ensuite capables d'atteindre la nourriture placée dans le labyrinthe aussi rapidement que d'autres rats qui ont eu cet apprentissage préalable avec nourriture comme récompenses. Or, lorsque l'apprentissage de ce labyrinthe se fait sans nourriture, il n'y a pas encore de changements observables dans le comportement des rats, le changement de comportement n'est donc pas nécessaire pour que l'apprentissage ait bien lieu. Le changement de comportement suite à une expérience n'est pas non plus suffisant pour considérer qu'il y ait de l'apprentissage. Par exemple, le changement de comportement suite à la perception d'un grand bruit n'est généralement pas considéré comme de l'apprentissage.

Pour résoudre certains de ces problèmes, d'autres définitions d'apprentissage ont été proposées. Ces définitions dites *mécanistes* reposent sur l'hypothèse que l'apprentissage implique un changement dans l'organisme et que ce changement est nécessaire mais pas suffisant pour observer un changement de comportement. Des exemples de ces définitions sont celles de Domjan "*l'apprentissage est un changement durable au niveau des mécanismes du comportement*" (DOMJAN, 2014) et Lachman "*l'apprentissage est un processus qui sous-tend le comportement qu'il ne faut pas confondre avec une conséquence de ce processus qu'est le changement de comportement*" (LACHMAN, 1997). Un avantage de la vision mécaniste de l'apprentissage est qu'elle permet d'expliquer qu'un changement de comportement n'est ni nécessaire ni suffisant pour inférer l'apprentissage. Étant donné que dans ce cas, l'apprentissage n'est qu'un mécanisme parmi d'autres facteurs provoquant le comportement, le changement de comportement n'est pas nécessaire car

l'impact de l'apprentissage peut être inhibé par d'autres facteurs. Le changement de comportement n'est pas non plus suffisant car d'autres facteurs peuvent en être responsable sans aucun apprentissage. De plus, la vision mécaniste permet de prendre en compte l'apprentissage latent en conservant l'effet de causalité contiguë. En effet, dans un premier temps, l'expérience résulte en un changement dans l'organisme. Ce changement permet de temporiser et d'exprimer le changement de comportement dans un second temps. Un premier problème pour ces définitions basées sur les mécanismes est qu'un changement dans l'organisme n'est pas simple à évaluer, et par conséquent le fait qu'il y ait eu apprentissage ou non est difficilement mesurable. Un autre problème pour ces définitions est qu'il est compliqué d'en sortir un moyen simple de définir ou différencier les sous-classes d'apprentissage (par exemple apprentissages non-associatifs, conditionnement classique, conditionnement opérant, etc.). Par conséquent les définitions mécanistes sont souvent complétées avec des définitions structurales ou fonctionnelles de sous-classes d'apprentissage.

Suite à ce constat sur les définitions historiques, les auteurs de DE HOUWER, BARNES-HOLMES et MOORS, 2013 repartent des idées de Skinner selon lesquelles le champ de la psychologie de l'apprentissage est l'étude de l'adaptation ontogénique. Une adaptation ontogénique est une adaptation de l'individu à son environnement qui se produit au cours de sa vie, là où une adaptation génétique est une adaptation à l'environnement à l'échelle d'une population ou d'une espèce. De Houwer et coauteurs définissent l'apprentissage comme les changements dans le comportement d'un organisme qui résultent des régularités dans l'environnement de cet organisme.

Cette définition se base sur trois concepts : (1) changement du comportement de l'organisme, (2) régularité dans l'environnement de l'organisme, (3) relation de causalité fonctionnelle entre (2) et (1), le changement du comportement est fonction de la régularité dans l'environnement de l'organisme.

Un comportement étant défini comme une réponse de l'organisme à l'environnement (dans le sens stimuli passés ou présents), le changement du comportement peut alors être défini comme une réponse différente de l'organisme dans une même situation environnementale (régularité).

Le concept de régularité est large. Est considéré comme régularité tout état de l'environnement de l'organisme qui implique un stimulus ou comportement unique répété plusieurs fois dans le temps, plusieurs stimuli ou comportements à un seul moment dans le temps, ou plusieurs stimuli ou comportements répétés plusieurs fois dans le temps. Le comportement de l'organisme fait partie intégrante de l'environnement et donc potentiellement des régularités menant au changement de comportement. Les états de l'environnement non réguliers sont ceux impliquant uniquement la présence d'un seul stimulus ou comportement à un seul moment dans le temps.

Enfin, De Houwer et coauteurs utilisent une causalité fonctionnelle entre comportement et environnement : le comportement est une fonction d'un élément de l'environnement. Étant donné qu'il s'agit de causes fonctionnelles et non pas mécaniques, il n'est pas nécessaire qu'il y ait contiguïté temporelle.

Notons aussi que la notion de stimulus est liée au contexte. Dans certaines expériences, le stimulus est un changement dans la perception d'un élément de l'environnement. Par exemple, les larves de poisson-zèbre (*Danio rerio*) réagissent à l'obscurité. Là où on imagine facilement un stimulus comme l'apparition d'un élément perceptible par l'organisme dans l'environnement comme de la lumière ou du son, l'exemple des poissons-zèbres rappelle qu'il peut aussi s'agir de

la disparition d'un élément dans l'environnement (en l'occurrence la lumière).

Une hiérarchie des apprentissages observés chez les animaux a été établie (MOORE, 2004 ; PERRY, BARRON et CHENG, 2013 ; SMITH-FERGUSON et BEEKMAN, 2019). Dans cette hiérarchie, les apprentissages les plus simples sont aussi les plus répandus dans le vivant. Les différents apprentissages (liste non exhaustive) sont regroupés dans les catégories suivantes, de la plus simple à la plus complexe :

- apprentissages non-associatifs :
  - habituation ;
  - sensibilisation spécifique au site ;
  - pseudo-conditionnement ;
- apprentissages associatifs :
  - conditionnement classique ;
  - apprentissage par inversion ;
  - décalage du pic ;
  - conditionnement opérant ;
- discriminations non-élémentaires :
  - apprentissage contextuel ;
  - apprentissage basé sur les nombres ;
  - apprentissage de la navigation ;
  - apprentissage de concepts ;
- apprentissages de niveau supérieur :
  - mimétisme et imitation ;
  - acquisition de langage.

Nous donnons ensuite les définitions accompagnées d'exemples pour les différents apprentissages indiqués dans la hiérarchie. Nous nous attardons surtout sur les apprentissages non-associatifs car ce sont eux que nous utiliserons comme inspiration dans la suite de cette thèse. En effet, ce sont des mécanismes qui nous paraissent suffisamment universels, au travers des espèces, pour qu'on puisse les considérer comme fondamentaux dans la capacité qu'a un organisme à acquérir de nouveaux comportements.

### **Apprentissages non-associatifs**

Les apprentissages non-associatifs (BYRNE et Robert D HAWKINS, 2015 ; ALEX et Catharine H RANKIN, 2017) correspondent à la modification du comportement ne nécessitant pas d'associations entre différents stimuli ou entre stimulus et réponse. L'organisme apprend à moduler ses réponses à partir d'un stimulus unique présenté seul ou de manière répétée : il y a une causalité entre l'apparition/la disparition du stimulus et l'observation du comportement. L'amplitude de la réponse peut être graduée et modifiée avec l'apprentissage. De manière simplifiée, on peut résumer les trois différents apprentissages non-associatifs de la manière suivante. Dans le premier apprentissage selon la hiérarchie (le plus simple), la modification du comportement est une réduction de la réponse déclenchée par le stimulus : on parlera alors d'*habituation*. Le second apprentissage, appelé *sensibilisation*, présente deux formes que l'on traite par la suite comme deux apprentissages différents. Soit il y a augmentation de la réponse déclenchée par le stimulus : il s'agit alors de *sensibilisation spécifique au site*. Soit il y a une généralisation de la réponse à un stimulus (par exemple sursauter après un bruit fort) comme réponse à d'autres stimuli auparavant neutres, qui ne provoquaient aucune réponse (sursaut face à un flash alors que ce dernier ne causait pas de réponse auparavant) : c'est le *pseudo-conditionnement*. Pour chacun

de ces apprentissages, nous allons ensuite donner la définition détaillée, puis l'illustrer à partir d'exemples basés sur différents organismes vivants.

## Habituation

*« Following Sharpless and Jasper's study [in 1956] great interest developed in habituation as a fundamental form of behavioral plasticity. As noted earlier, it occurs for virtually all behavioral responses in virtually all organisms. » THOMPSON, 2009*

L'habituation, aussi appelée accoutumance (MÉDIONI et ROBERT, 1969), est l'apprentissage non-associatif reconnu comme le plus simple et le plus répandu dans le vivant. La caractéristique essentielle de l'habituation est la diminution de la réponse de l'organisme à un stimulus suite à la répétition ou prolongation de ce même stimulus (MOORE, 2004). Pour être de l'habituation, cette diminution doit être non attribuable à de la fatigue ou à de l'adaptation sensorielle. Lors des expériences cherchant à identifier l'habituation, un autre stimulus provoquant la même réponse chez l'organisme peut être utilisé pour exclure ces autres explications possibles de la diminution. Si nous faisons un exemple théorique avec deux stimuli d'une même modalité sensorielle  $S1$  et  $S2$  (imaginons que ce soit des sons), et un troisième stimulus d'une autre modalité  $S3$  (un flash de lumière). Par défaut, notre organisme théorique répond aux trois stimuli avec une réponse identique : il recule d'une certaine distance. Au fur et à mesure des répétitions du stimulus  $S1$ , l'organisme recule de moins en moins, jusqu'à ne plus reculer du tout. S'arrêter ici ne permet pas de conclure qu'il y a habituation, l'organisme pourrait ne plus avoir assez d'énergie pour reculer. Pour tester si c'est de la fatigue, il faut évaluer sa réponse au stimulus  $S3$  : s'il ne recule pas, il s'agirait effectivement plutôt de fatigue que d'habituation. S'il recule, avant de trancher pour l'habituation, il faut tester l'adaptation sensorielle avec le stimulus  $S2$  (de même modalité sensorielle que  $S1$  sur lequel l'organisme a été "entraîné"). À nouveau, on observe sa réponse à ce stimulus : soit il ne recule pas, et il s'agirait alors plutôt d'adaptation sensorielle, soit il recule, dans ce cas il s'agirait d'habituation.

Richard F. Thompson a publié un historique de l'habituation (THOMPSON, 2009). Le phénomène est connu depuis l'antiquité, et a connu un intérêt particulièrement marqué de la fin du XIX<sup>e</sup> siècle au début du XX<sup>e</sup> siècle. Les expériences et observations de cette période ont mené à une synthèse dans THOMPSON et SPENCER, 1966, où neuf caractéristiques de l'habituation sont identifiées et décrites. Cette description précise de l'habituation a ensuite été revue et complétée par Catharine H RANKIN et al., 2009, et nous l'avons retranscrite dans la table 2.1. La caractéristique fondamentale de l'habituation est la diminution progressive de l'un ou plusieurs paramètres d'une réponse à un stimulus suite à la présentation répétée de celui-ci. D'autres caractéristiques permettent de distinguer l'habituation d'autres sources pouvant expliquer la diminution de la réponse (fatigue ou adaptation sensorielle par exemple) : la récupération spontanée lorsqu'elle est dépendante de la fréquence du stimulus, la spécificité au stimulus et la déshabituation. La récupération spontanée se caractérise par la récupération partielle ou totale de cette réponse une fois la stimulation arrêtée, elle est d'autant plus rapide que la fréquence du stimulus a été haute pendant la phase d'habituation. La spécificité au stimulus est la caractéristique utilisée dans l'exemple théorique précédent (avec  $S1$ ,  $S2$  et  $S3$ ) pour illustrer une façon de dissocier l'habituation de la fatigue et de l'adaptation sensorielle. La déshabituation se caractérise par une récupération plus ou moins importante et instantanée de la réponse à la suite de la présentation d'un stimulus différent de celui utilisé pour l'habituation. Cette dépendance à un autre stimulus implique que la déshabituation peut être sujette elle-même à l'habituation si ce stimulus est lui aussi répété, ce qui conduit à une récupération de plus en plus faible de la réponse initiale.

Certaines propriétés du stimulus peuvent renforcer l'habituation : un stimulus de faible intensité et/ou présenté à haute fréquence permet une habituation plus rapide ou plus prononcée. La répétition du stimulus après que la diminution de la réponse soit devenue nulle peut avoir d'autres effets, comme par exemple un début plus tardif de la récupération spontanée une fois la stimulation finie. Répéter plusieurs phases d'habituation après récupération complète de la réponse induit une habituation plus rapide ou plus prononcée, il est question de potentialisation de l'habituation. Enfin, en fonction du protocole de stimulation, il a été remarqué que certaines des propriétés de la diminution de la réponse peuvent durer entre plusieurs heures à plusieurs semaines. Par exemple, lors d'une seconde séance d'habituation, la diminution de la réponse peut être plus rapide, la réponse initiale peut être plus basse, la moyenne des réponses peut être plus basse, ou encore les réponses peuvent être moins fréquentes.

Un des organismes les plus simples où l'habituation a été observée est *Physarum polycephalum* ((BOISSEAU, D. VOGEL et DUSSUTOUR, 2016)). De manière générale et simplifiée, le cycle de vie de ce myxomycète est séparé en deux parties en fonction de son matériel génétique : *haploïde* ou *diploïde* (2.1). Dans chacun des cas, il a une forme mobile : l'*amibe* et le *plasmode*. La phase où l'habituation a été observée est sa forme plasmode, récemment surnommée le "blob". Lorsqu'il est dans cette forme, c'est un organisme monocellulaire contenant plusieurs noyaux, il ne possède donc aucun neurone. Cependant, malgré cette relative simplicité, il est capable de se déplacer par extension et rétraction, et a besoin d'apprendre pour survivre. Dans les expériences de l'article, nous pouvons voir qu'il se propage plus ou moins rapidement à travers le substrat pour atteindre de la nourriture. En particulier, la quinine et la caféine ont un effet répulsif sur le blob, effet mesuré par le temps qu'il met à se rapprocher du substrat, à le traverser, mais aussi la forme et la taille de son extension (*pseudopode*) traversant le substrat. La répétition de la présentation du répulsif montre que le blob est capable d'habituation, que ce soit en rejoignant ou en traversant le substrat de façon plus rapide, mais aussi en utilisant un pseudopode plus grand, au fur et à mesure des répétitions.

Loin d'être anecdotique, l'habituation semble au contraire être fondamentale pour les êtres vivants. En effet, elle semble omniprésente dans la phylogénie, représentée sous différents mécanismes ayant évolué indépendamment (E. EISENSTEIN, D. EISENSTEIN et SMITH, 2001 ; MOORE, 2004 ; THOMPSON, 2009 ; PERRY, BARRON et CHENG, 2013). *A minima*, un organisme doté d'habituation est capable de diminuer progressivement un réflexe défensif (par exemple un escargot se repliant dans sa coquille) face à un stimulus sans conséquences dangereuses et cela sans avoir l'attrait permettant des apprentissages plus complexes pour faire le lien entre le stimulus et la non-dangerosité qui suit. Cet organisme peut ainsi reprendre son activité en cours (par exemple chercher sa nourriture). Sans cela, on pourrait imaginer l'organisme bloqué tant que le stimulus est présent, ce qui lui coûterait l'énergie de son comportement de réponse ainsi que la perte de temps où il ne cherche plus sa nourriture. Apprendre à considérer un stimulus récurrent comme du bruit à ignorer permet aussi à l'organisme de se concentrer sur d'autres stimuli éventuellement plus significatifs pour lui. Cependant, la question des avantages adaptatifs de l'habituation reste ouverte (RANDLETT et al., 2019 ; MCDIARMID, YU et Catharine H RANKIN, 2019), ils pourraient aller au delà de permettre à l'organisme d'ignorer les stimuli non significatifs pour en plus faciliter les changements de stratégie comportementale. Cette hypothèse repose sur l'observation de l'habituation à certains stimuli nocifs (donc significatifs) et sur la multiplicité des mécanismes de l'habituation chez *Caenorhabditis elegans* et chez les larves de poissons-zèbres *Danio rerio*.

1. **Habituation** : l'application répétée d'un stimulus provoque une diminution progressive de l'un ou plusieurs paramètres d'une réponse jusqu'à un niveau asymptotique.
2. **Récupération spontanée** : si la stimulation est arrêtée, la réponse récupère au moins partiellement.
3. **Potentialisation de l'habituation** : alterner de manière répétée des sessions d'habituation et de récupération spontanée conduit à une habituation progressivement plus rapide et/ou plus prononcée.
4. **Effet de l'intervalle interstimulus** : étant données des propriétés de stimulus égales, la stimulation à haute fréquence entraîne une diminution plus rapide et/ou plus prononcée de la réponse et une récupération spontanée plus rapide.
5. **Effet de l'intensité du stimulus** : plus le stimulus est de faible intensité, plus la diminution de la réponse est rapide et/ou prononcée (en mesures relatives et non absolues de la réponse). Les stimuli très intenses peuvent ne pas présenter de diminution de la réponse.
6. **Habituation en dessous de zéro** : des effets de la stimulation peuvent continuer à s'accumuler même après que la réponse ait été diminuée à un niveau asymptotique ou nul (absence de réponse), comme par exemple en retardant le début de la récupération spontanée.
7. **Spécificité au stimulus** : la diminution à la réponse montre une certaine spécificité du stimulus.
8. **Déshabituation** : si un stimulus différent est inséré dans une série de stimuli provoquant de l'habituation, la réponse au stimulus original se voit renforcée.
9. **Habituation de la déshabituation** : pour une réponse à un stimulus donné, répéter habituation et déshabituation conduit à la diminution de l'ampleur de la déshabituation (c'est-à-dire que le renforcement de la réponse est de plus en plus faible).
10. **Habituation à long terme** : selon le protocole de répétition du stimulus, les modifications des propriétés de la réponse peuvent durer des heures, des jours ou des semaines. (par exemple, une réhabilitation plus rapide que la ligne de base, des réponses initiales plus petites que la ligne de base, des réponses moyennes plus petites que la ligne de base, des réponses moins fréquentes que la ligne de base).

TABLE 2.1 – Caractéristiques de l'habituation (MCDIARMID, YU et Catharine H RANKIN, 2019; Catharine H RANKIN et al., 2009)

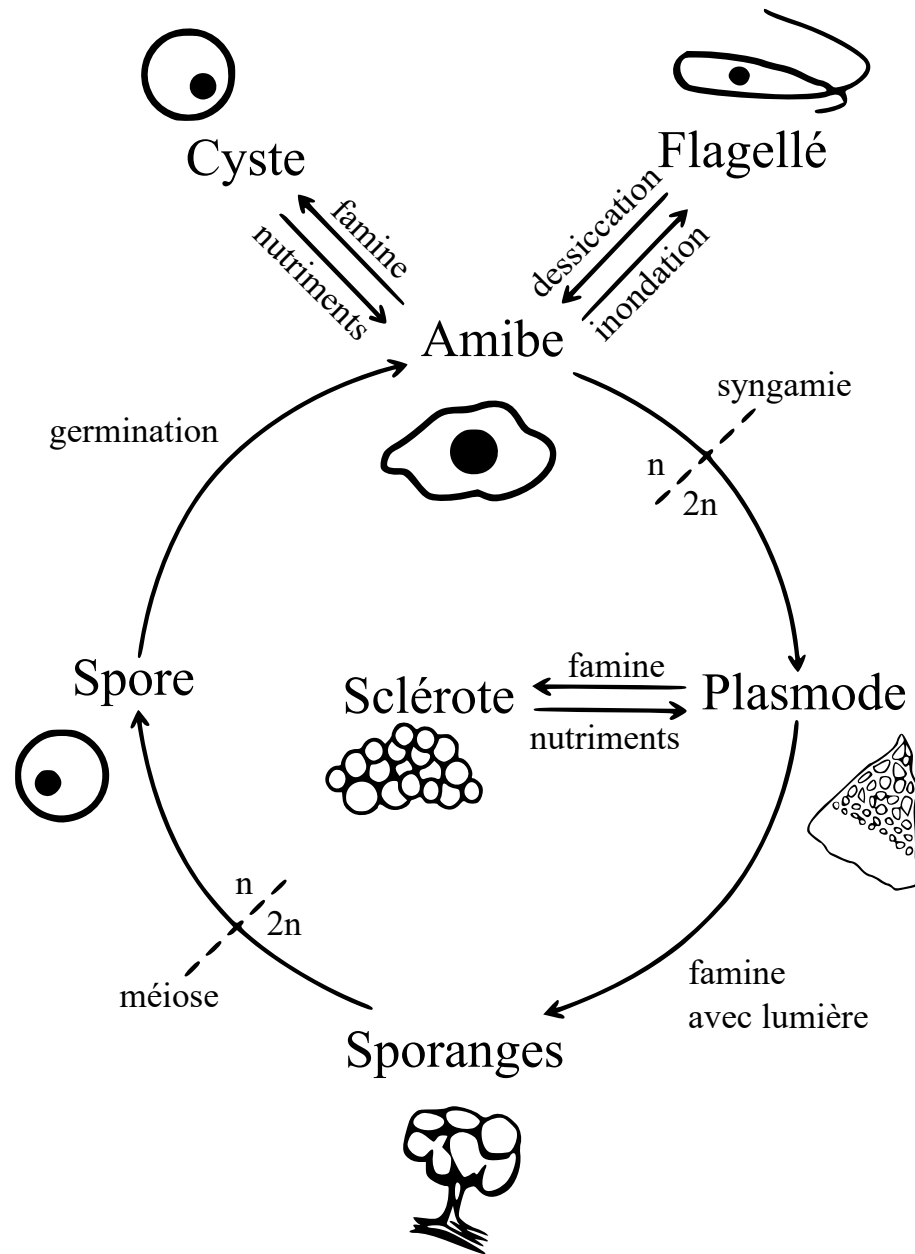


FIGURE 2.1 – Cycle de vie de *Physarum polycephalum* (figure adaptée de BARRANTES, LEIPZIG et MARWAN, 2012 et corrigée en suivant HARA et OHTA, 1981). Ses deux phases mobiles sont l'amibe (avec un noyau haploïde) et le plasmode (avec plusieurs noyaux diploïdes). La fusion de deux amibes (syngamie) forme un zygote qui donne un plasmode après division de son noyau. En cas de manque de nutriments et de présence de lumière, le plasmode va se couvrir de sporanges. Arrivées à maturation, ces sporanges relâchent les spores, qui libèrent les amibes après germination.

**Sensibilisation** La sensibilisation est l'autre apprentissage non-associatif, lui aussi très représenté dans le vivant. Définie comme le renforcement de la réponse d'un organisme à un stimulus après présentation répétée de ce stimulus, elle englobe deux formes (BYRNE et Robert D HAWKINS, 2015) : sensibilisation spécifique au site et pseudo-conditionnement. Selon les travaux, chacune de ces formes peut être appelée directement sensibilisation malgré leurs différences, ce qui est source de confusion.

**Sensibilisation spécifique au site** La sensibilisation spécifique au site est définie en tant qu'augmentation de la réponse comportementale après la présentation répétée du même stimulus, effet opposé de l'habituation. De plus, l'effet de l'intensité du stimulus est aussi l'opposé de ce que l'on observe pour l'habituation puisque plus le stimulus est intense et plus l'augmentation de la réponse est marquée, alors qu'un stimulus faible ne provoquera pas (ou peu) de sensibilisation spécifique au site. Similairement à l'habituation, la sensibilisation spécifique au site est un phénomène avec récupération spontanée et elle peut être observée sur plus ou moins long terme selon les conditions de répétition du stimulus.

Par exemple, les expériences de GROVES, LEE et THOMPSON, 1969 étudient l'habituation et la sensibilisation spécifique au site chez le chat spinal (avec moelle épinière sectionnée à la douzième vertèbre). La réponse évaluée est le réflexe de flexion, réponse déclenchée par stimulation électrique des zones distales de la peau des membres postérieurs ou par stimulation directe du nerf péronier superficiel (résultats comparables dans les deux cas). Leurs résultats indiquent que la sensibilisation spécifique au site à court terme induit une réponse plus forte en fonction de la fréquence du stimulus. La figure 2.2 illustre les différences entre la sensibilisation spécifique au site à court terme et celle à long terme. À court terme, elle apparaît très rapidement (le pic est atteint au premier bloc de cinq réponses) mais s'estompe aussi très rapidement (retour quasiment à la réponse témoin vers la quinzième réponse). L'effet à long terme est plus progressif, il n'atteint un niveau asymptotique qu'au delà de la centième réponse à un niveau supérieur à l'effet à court terme. Enfin, sur la figure 2.3, nous pouvons voir que la sensibilisation spécifique au site à long terme et l'habituation peuvent s'observer en même temps. L'habituation s'exprime davantage à faible intensité, proportionnellement à la fréquence. La sensibilisation spécifique au site est, à l'inverse, prédominante à forte intensité et à basse fréquence. Nous pouvons noter que sur la figure 2.2, l'effet de l'habituation en parallèle de la sensibilisation est dur à évaluer. Pourtant, selon les autres éléments de l'article, la fréquence est assez haute (16 pulsations par seconde, soit le double de la plus haute fréquence de la figure 2.3) et l'intensité assez basse (pulsations de 10V) pour que l'habituation soit significative.

Dans les expériences de Catherine H RANKIN, BECK et CHIBA, 1990, les auteurs observent de la sensibilisation spécifique au site à court terme chez le ver *Caenorhabditis elegans*. Contrairement aux expériences de GROVES, LEE et THOMPSON, 1969 sur le chat spinal, la sensibilisation est provoquée par une brusque augmentation de la fréquence du stimulus. Le stimulus utilisé est un tapotement sur le côté de la boîte de Pétri contenant un ver. Après un tapotement, ce dernier recule d'une certaine distance. La moyenne des distances reculées par plusieurs vers permet d'évaluer l'habituation et la sensibilisation spécifique au site. Le protocole pour la sensibilisation est le suivant : il y a cinq stimuli, chacun séparé de deux minutes. Les deux premiers sont un seul tapotement, et leur moyenne sert de référence (réponse témoin). Ensuite, le troisième stimulus est un enchaînement de six tapotements, puis les deux derniers sont à nouveau des tapotements uniques (respectivement le test 1 et le test 2). Lors du test 1 les vers montrent une importante sensibilisation (ils reculent en moyenne du double de la réponse témoin) alors que pour le test 2



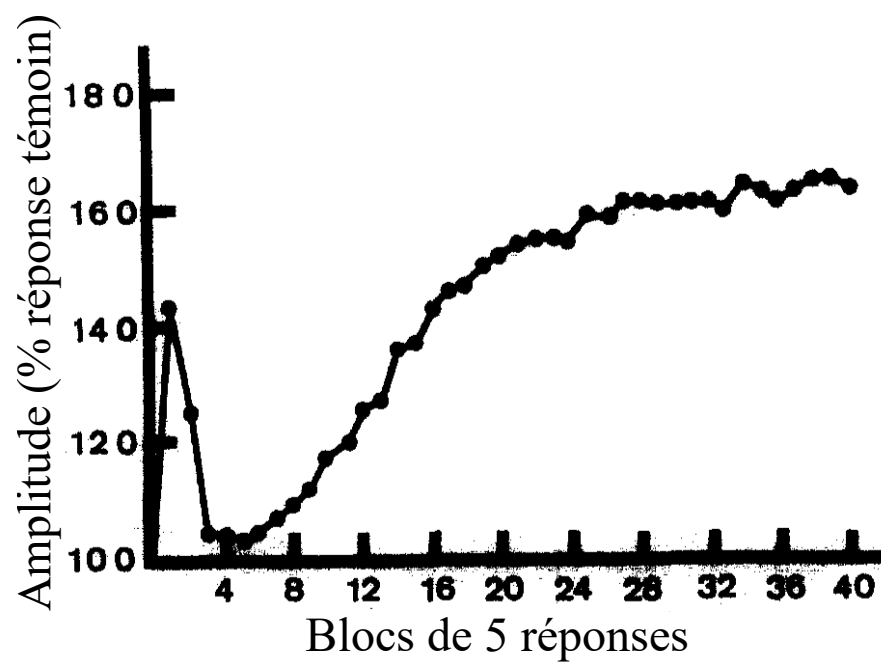


FIGURE 2.2 – Effet des sensibilisations spécifiques au site court terme et long terme (figure 6 de GROVES, LEE et THOMPSON, 1969), avec une fréquence de 16 stimuli par seconde et avec une intensité de 10V. La sensibilisation spécifique au site à court terme est relativement éphémère, le pic est atteint en 5 réponses et la réponse la plus haute vaut 140% de la réponse témoin. L'effet s'estompe vers la 15<sup>e</sup> réponse. La sensibilisation spécifique au site à long terme est plus lente et progressive à se mettre en place, elle commence à la 25<sup>e</sup> réponse et atteint un niveau asymptotique d'environ 160% de la réponse témoin après la 120<sup>e</sup> réponse.

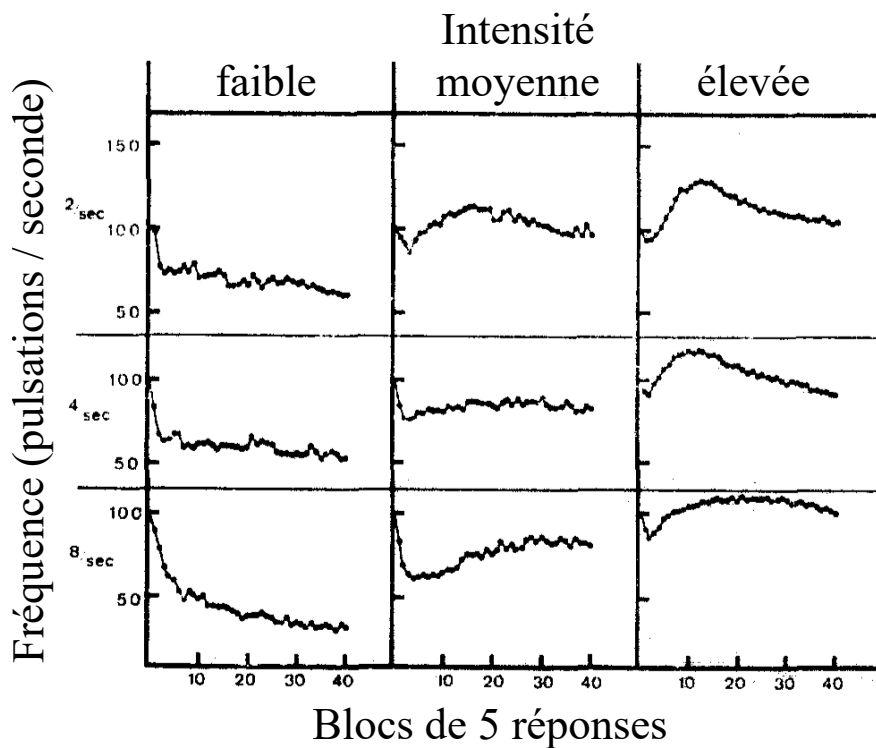


FIGURE 2.3 – Effets de l'intensité et de la fréquence du stimulus sur l'habituation et la sensibilisation spécifique au site (figure 7 de GROVES, LEE et THOMPSON, 1969). L'habituation est prédominante à faible intensité, et est plus marquée proportionnellement à la fréquence. La sensibilisation spécifique au site est, à l'inverse, prédominante à forte intensité et à basse fréquence.

l'effet est déjà estompé (ils reculent de la même distance que la réponse témoin).

**Pseudo-conditionnement** Le pseudo-conditionnement est une forme de sensibilisation permettant à une réponse comportementale inconditionnelle *RI* à un stimulus *inconditionnel SI* donné de se généraliser, temporairement, comme réponse "conditionnelle" *RC* à un autre stimulus *conditionnel SC* (Robert D HAWKINS et KANDEL, 1984). La différence avec le *conditionnement classique*, qui est un *apprentissage associatif*, est que dans le cas du pseudo-conditionnement la réponse conditionnelle *RC* a lieu sans avoir eu besoin de présenter les deux stimuli *SI* et *SC* ensemble, au même moment. Là où le conditionnement classique permet de généraliser la réponse inconditionnelle *RI* à *SC* uniquement, via son association avec le *SI*, le pseudo-conditionnement permet de la généraliser à d'autres stimuli sans associations.

Le lièvre de mer *Aplysia californica* possède un réflexe défensif lui permettant de rétracter ses branchies et/ou son siphon. Dans les expériences de CAREW, WALTERS et KANDEL, 1981, ce réflexe permet la démonstration du pseudo-conditionnement (appelé "sensibilisation" dans l'article) chez le lièvre de mer. Le stimulus inconditionnel *SI* utilisé est un choc électrique au niveau de la queue de l'organisme, ce qui provoque comme réponse inconditionnelle *RI* une rétractation importante des branchies et du siphon. Le stimulus conditionnel *SC* utilisé est le contact d'une brosse en nylon sur le siphon, déclenchant comme réponse *naïve RN* (réponse avant apprentissage) une légère rétractation du siphon uniquement. Dans les expériences concernant le pseudo-conditionnement, seule la durée de rétractation du siphon est mesurée. Dans une expérience, des groupes de lièvres de mer sont testés en fonction du nombre de présentation du *SI* et du temps avant présentation du *SC*. Le *SI* est présenté soit une unique fois, soit quinze fois, puis les lièvres de mer sont ensuite stimulés par le *SC* après cinq, trente ou soixante minutes. Les résultats sont qu'une seule présentation du *SC* est suffisante pour déclencher le pseudo-conditionnement (donc une durée de rétractation du siphon significativement plus longue que lors de la réponse *RN*) même si son effet est plus marqué après quinze présentations du *SI*. De plus, le pseudo-conditionnement n'apparaît pas dans les groupes testés après cinq minutes, mais dans les groupes testés à trente ou soixante minutes. Une autre de leur expérience montre que les effets du pseudo-conditionnement sont aussi observés après vingt-quatre heures. Dans cet exemple, l'effet de généralisation est cependant à nuancer : la réponse initiale au *SC* n'est pas neutre mais un léger retrait du siphon, et seul le retrait du siphon est mesuré pour ce protocole donc on ne sait pas si le retrait des branchies est lui généralisé ou non.

### Apprentissages associatifs

Les apprentissages associatifs sont ceux où il y a apprentissage d'une association. Cette association peut être entre deux stimuli, comme dans le *conditionnement classique*, et permet alors de transmettre la réponse de l'un vers l'autre. Elle peut aussi être entre les stimuli, le comportement et les conséquences (positives ou négatives pour l'organisme) de ce comportement : on parle alors de *conditionnement opérant*.

**Conditionnement classique (ou Pavlovien)** Le conditionnement classique, aussi appelé conditionnement Pavlovien, est un apprentissage où une réponse inconditionnelle *RI* à un stimulus inconditionnel *SI* devient la réponse conditionnelle *RC* d'un autre stimulus, appelé stimulus conditionnel *SC*, généralement neutre (il ne provoque aucune réponse spécifique ou significative avant l'apprentissage). Pour qu'il y ait conditionnement classique, le *SI* et le *SC* doivent être présentés en même temps plusieurs fois pour que l'association entre *RI* et *SC* se mette en place

pour produire *RC* lors des prochaines présentations de *SC*. L'exemple classique pour illustrer cet apprentissage est avec le chien de Pavlov, la nourriture comme *SI*, la salivation comme *RI* et un son de cloche comme *SC*. Initialement, la cloche ne provoque pas de salivation. Mais si la cloche et la nourriture sont présentées ensemble au chien de manière répétées, faire tinter la cloche par la suite provoque la salivation du chien. Pour discriminer le conditionnement classique des deux formes de sensibilisation, il est nécessaire d'avoir des groupes de contrôle spécifiques, et que le groupe testé sur le conditionnement classique montre une réponse conditionnée plus importante. Le groupe de contrôle pour la sensibilisation spécifique au site ne sera stimulé que par le *SC*, là où le groupe de contrôle pour le pseudo-conditionnement ne sera stimulé que par le *SI*. D'autres groupes de contrôle supplémentaires sont parfois mis en place : *SC* et *SI* explicitement non appariés ou *SC* et *SI* présentés dans une séquence aléatoire par exemple (CAREW, WALTERS et KANDEL, 1981 ; DUSSUTOUR, 2021). Si nous reprenons l'exemple du lièvre de mer *Aplysia californica* dans le même article (CAREW, WALTERS et KANDEL, 1981), dans une autre expérience un groupe est testé avec la présentation des deux stimuli *SI* et *SC* en même temps pour mesurer l'effet du conditionnement classique face à d'autres groupes : naïf (pas de stimuli), *SI* seul, *SC* seul, *SC* et *SI* non appariés ou *SC* et *SI* présentés aléatoirement. La présentation des stimuli (en fonction des groupes) a lieu toutes les cinq minutes, pour un total de trente-et-une stimulations. Immédiatement après la première stimulation et ensuite après toutes les cinquièmes stimulations le *SC* seul est présenté à tous les groupes pour évaluer l'évolution de la durée de rétractation du siphon. Une dernière stimulation, *SC* seul, a lieu le lendemain. Dans le groupe où *SI* et *SC* sont appariés, la durée de rétractation du siphon en réponse au *SC* seul est de plus en plus longue au fur et à mesure des stimulations, alors que pour tous les autres groupes il n'y a pas d'augmentation significative. Par contre, après vingt-quatre heures, le groupe "*SI* seul" répond à la stimulation avec une durée de rétractation accrue, même si moindre que celle du groupe appariés. Cela illustre deux points : chez le lièvre de mer, l'effet du conditionnement classique est plus important que celui du pseudo-conditionnement, et est observable plus rapidement.

**Apprentissage par inversion** L'apprentissage par inversion ("*reversal learning*") est un cas particulier de conditionnement classique. Il s'agit d'apprendre à inverser une (ou plusieurs) association(s) (PERRY, BARRON et CHENG, 2013). Par exemple, si un organisme a déjà appris les associations bleu/nourriture et blanc/choc, il est capable d'apprentissage par inversion s'il arrive à apprendre à la place bleu/choc et blanc/nourriture. Cela nécessite la plasticité d'oublier des associations pour les remplacer par d'autres.

**Décalage du pic** Le décalage du pic ("*peak shift*") intervient lorsqu'un organisme pré-entraîné sur deux stimuli d'une même modalité (positivement sur *S+* et négativement sur *S-*) répond davantage à un stimulus proche de *S+* mais encore plus loin de *S-* (PERRY, BARRON et CHENG, 2013). Par exemple, si les stimuli sont de la lumière, avec *S+* une longueur d'onde de 600nm (un jaune) et *S-* une longueur d'onde 500nm (un bleu ciel), lorsque l'on présente à l'organisme différentes couleurs, il répondra le plus pour une longueur d'onde plus grande encore que *S+* (disons 620nm, un jaune orangé).

### Conditionnement opérant (ou instrumental)

*« In operant conditioning we "strengthen" an operant in the sense of making a response more probable or, in actual fact, more frequent. In Pavlovian or "respondent" conditioning we simply increase the magnitude of the response elicited by the conditioned stimulus and shorten the time which elapses between stimulus and response. »*

Burrhus Frederic SKINNER, 1965

Le conditionnement opérant est aussi un apprentissage associatif, mais l'association n'est plus entre différents stimuli mais entre un comportement de l'organisme et les conséquences qui en résultent.

Cet apprentissage a été caractérisé par B.F. Skinner (B. SKINNER, 1938; Burrhus Frederic SKINNER, 1965), dont les travaux sont basés sur la *loi de l'effet* d'E.L. Thorndike (THORNDIKE, 1898). La loi de l'effet propose que "*les réponses qui produisent un effet satisfaisant dans une situation donnée sont plus susceptibles de se reproduire dans cette situation, et les réponses qui produisent un effet désagréable sont moins susceptibles de se reproduire dans cette situation.*" (GRAY, 2011).

Skinner approfondie la loi de l'effet avec plusieurs notions : celle d'opérants qui peuvent être neutres, renforçateurs ou punisseurs, ainsi que celles de renforcement et de punition. Les opérants sont vus comme des réponses de l'environnement au comportement de l'organisme, ils sont la conséquence à laquelle le comportement est associé lors de l'apprentissage. Un opérant est neutre s'il n'a pas d'impact sur la probabilité que le comportement soit répété, renforçateur s'il augmente cette probabilité ou punisseur s'il la diminue. Un renforçateur peut être positif (par exemple de la nourriture que le comportement permet d'obtenir) ou négatif (par exemple un choc que le comportement permet d'éviter). De même, un punisseur peut être positif (par exemple un choc que le comportement provoque) ou négatif (par exemple fermer l'accès à la nourriture suite au comportement). Il y a renforcement dans les cas d'apprentissage où l'opérant est un renforçateur et punition dans les cas où l'opérant est un punisseur.

### Discriminations non-élémentaires

Les discriminations non-élémentaires sont les différentes formes d'apprentissages qui nécessitent plus d'éléments à apprendre que dans les apprentissages non-associatifs ou associatifs, ou qui reposent sur une combinaison de plusieurs apprentissages (conditionnements classique et opérant par exemple). Cependant, cette classification reste subjective car l'évaluation de la complexité des tâches à effectuer est basée sur l'intuition, plutôt que sur les mécanismes nécessaires par exemple (PERRY, BARRON et CHENG, 2013).

**Apprentissage contextuel** L'apprentissage contextuel permet de discriminer des associations stimulus/récompense en fonction de contexte (autres stimuli pouvant se traduire comme une luminosité différente, lieux différents, odeurs différentes, directions de vol différentes etc.). L'organisme peut ainsi apprendre qu'un stimulus  $A$  perçu dans un contexte  $C_1$  est positif, alors qu'un autre stimulus  $B$  ne l'est pas ( $AC_1+$  et  $BC_1-$ ), et apprendre l'inverse dans un contexte  $C_2$  ( $AC_2-$  et  $BC_2+$ ). Cet apprentissage est considéré plus complexe qu'un "simple" apprentissage associatif car l'organisme doit pouvoir associer plus qu'une conséquence à un même stimulus.

**Apprentissage basé sur les nombres** Un organisme capable de discrimination basée sur les nombres arrive à faire l'association entre un choix où la seule différence est le nombre d'éléments et la conséquence de ce choix (par exemple, obtenir de la nourriture). Dans les expériences de GROSS et al., 2009, avec un labyrinthe en "Y", les abeilles apprennent dans un premier temps que la nourriture est derrière la porte avec un certain nombre d'éléments et pas l'autre (par exemple 2 contre 3 points bleus). Ensuite, elles choisissent de manière significative la porte avec le bon nombre d'éléments, même en cas de changement de symboles.

**Apprentissage de la navigation** Dans PERRY, BARRON et CHENG, 2013, la navigation est décrite comme une forme complexe d'apprentissage impliquant une combinaison de conditionnement classique de repères (position du soleil, des étoiles, panorama de la ligne d'horizon, etc.) et de conditionnement opérant d'indices proprioceptifs (par exemple le comptage des pas pour l'intégration de chemin). Il semble utile à tout organisme mobile ayant un abri fixe (tanière ou autre) d'avoir un moyen de navigation lui permettant de retrouver son abri après être parti chercher de la nourriture par exemple.

**Apprentissage de concepts** L'apprentissage de concepts inclut tout apprentissage de relations relatives entre objets (comme par exemple similaire, différent, au dessus, en dessous, etc.). Les abeilles *Apis mellifera* ont montré qu'elles pouvaient apprendre le concept de dessus/dessous via un labyrinthe en "Y" (AVARGUÈS-WEBER, DYER et GIURFA, 2011). Les seules informations constantes lors des différents essais sont des stimuli visuels où un référent (barre horizontale, croix ou disque) est toujours présent et une cible est soit au dessus du référent pour l'un des bras du "Y" soit en dessous pour l'autre bras, avec une récompense (du sucre) au bout du "bon" chemin. Après apprentissage, les abeilles sont testées sur des cibles qu'elles n'ont jamais vues auparavant et parviennent à obtenir la récompense.

### Apprentissages de niveau supérieur

**Imitation** L'imitation est un apprentissage nécessitant que l'organisme soit capable de copier une nouvelle réponse (non instinctive ou non typiques de l'espèce), sans recevoir de récompense ou autre forme de renforcement, et ce dans des situations excluant tout autre explication plus simple (MOORE, 2004).

En plus des humains, l'imitation a été observée chez les cétacés (dauphins), les oiseaux (perroquets gris d'Afrique) et les grands singes (chimpanzés). Dans l'article de TAYLER et SAAYMAN, 1973, un dauphin en captivité imite le comportement d'un plongeur qui nettoie régulièrement les algues d'une fenêtre sous-marine. Lors du nettoyage, le plongeur se maintient d'une main au cadre en acier de la fenêtre, et enlève les algues entre le cadre et la vitre à l'aide d'un pinceau dans son autre main. Lorsque le plongeur est absent, le dauphin se maintient au cadre de la fenêtre avec une de ses nageoires et nettoie avec des mouvements similaires la jonction entre cadre et vitre à l'aide de ce qu'il trouve (plume de mouette, pierres, papier, etc.). De plus, le dauphin imite aussi le son de la valve de plongée, tout en relâchant des flux de bulles d'une manière similaire à celui s'échappant de l'appareil de plongée.

**Acquisition de langage** L'apprentissage de langage est le plus complexe de cette hiérarchie. Selon MOORE, 2004, pour acquérir un langage gestuel ou vocal, il est nécessaire d'être capable de combiner plusieurs autres processus d'apprentissage comme par exemple la mimique vocale, la phonétique, l'apprentissage de discriminations, les associations entre stimuli neutres, l'apprentissage d'habiletés (en particulier d'outils), la formation de concepts abstraits, l'apprentissage moteur guidé (*putting through*), l'imitation de mouvements et la pantomime. Un exemple d'apprentissage moteur guidé est lorsque qu'une personne indique à l'apprenant une certaine posture en dirigeant directement ses membres (comme dans certains sports, utilisations de certains instruments de musiques, etc.).

Dans cette thèse, nous avons choisi de nous focaliser sur les apprentissages non-associatifs (BYRNE et Robert D HAWKINS, 2015; ALEX et Catharine H RANKIN, 2017). Ce choix repose

sur deux arguments. Le premier est le fait que ce sont les apprentissages les plus répandus dans le vivant, simples mais possiblement fondamentaux, avec sans doute la capacité d'induire des changements de comportement utiles ou même nécessaires pour faire face à ceux de l'environnement. De plus, ils pourraient être constitutifs d'apprentissage de niveaux plus élevés (conditionnement classique par exemple), ou nécessaires pour leur mises en œuvre ou leur régulation. Le second est notre approche *bottom-up* de vouloir un agent capable d'augmenter son répertoire comportemental de manière incrémentale avec les outils les plus simples. L'exploration de l'état de l'art des modèles qui suit, qu'ils cherchent à reproduire des résultats observés en biologie ou à être appliqués en pour des robots ou des agents artificiels, est par conséquent aussi limité à ces apprentissages.

## 2.2 Agents artificiels utilisant de l'apprentissage non-associatif

Plusieurs travaux de modélisation utilisent les apprentissages non-associatifs, certains de manière conceptuelle et d'autres de manière plus opérationnelle, dans des domaines variés (psychologie, médecine, intelligence artificielle ou apprentissage automatique). Tous les travaux référencés ici traitent de l'habituation, avec parfois la sensibilisation spécifique au site ou le pseudo-conditionnement en plus, dans les deux cas sous le terme de sensibilisation.

Nous parlons dans un premier temps de modèles conceptuels où les auteurs cherchent à retrouver des résultats observés en biologie ou à simuler les différentes caractéristiques de l'apprentissage sur lequel ils travaillent. Ensuite, nous référençons plusieurs travaux qui se sont concentrés sur l'utilisation de l'habituation en tant qu'outil de détection de nouveauté. Les derniers modèles concernent diverses applications des apprentissages non-associatifs (amélioration des performances de réseaux de neurones ou de processeurs neuromorphiques, système d'attention, adaptation à l'environnement en robotique).

### 2.2.1 Modèles conceptuels

Ces premiers travaux de psychologie proposent des modèles permettant de retrouver des résultats observés lors d'expérimentations sur le vivant pour appuyer des théories sur les mécanismes biologiques responsables de l'apprentissage.

Dans le modèle de conditionnement classique de Rescorla-Wagner (RESCORLA, 1972), l'évolution de l'association entre  $SI$  et  $SC$  dépend de la prédiction du  $SI$ . L'observation sur laquelle cette caractéristique repose est qu'un  $SI$  signalé, donc attendu, est moins propice à l'établissement du conditionnement. Le parallèle qu'a fait Wagner avec l'habituation est que la répétition d'un stimulus dans un même contexte devient prévisible par association, et cela a conduit au modèle conceptuel *Sometimes Opponent Processes* (ou *SOP*, A. WAGNER, 1981 ; E. H. VOGEL, PONCE et A. R. WAGNER, 2019). Que l'habituation soit dépendante du contexte est contesté car peu d'observations soutiennent cette théorie (HALL et RODRÍGUEZ, 2020). Dans le modèle SOP, un stimulus  $S$  est représenté comme un ensemble d'éléments, dont chacun peut alterner entre trois états : inactif  $I_S$ , activité primaire 1  $A1_S$ , activité primaire 2  $A2_S$ . Au départ, tous les éléments sont dans l'état inactif  $I_S$ . Lorsque  $S$  est présenté, chaque élément a la probabilité  $p1_S$  de passer à l'état  $A1_S$ . Puis à chaque pas de temps les éléments dans l'état  $A1_S$  ont la probabilité  $pd1_S$  de décliner dans l'état  $A2_S$ , et ceux dans l'état  $A2_S$  ont la probabilité  $pd2_S$  de retourner dans l'état inactif  $I_S$ . La réponse  $R$  au stimulus est fonction de la proportion d'éléments dans l'état  $A1_S$  pondérés par un poids  $w1$  et de la proportion d'éléments dans  $A2_S$  qui sont eux

pondérés par un autre poids  $w_2$  :

$$R = f \times (w_1 \times PA_{1S} + w_2 \times PA_{2S}) \quad (2.1)$$

avec  $f$  une fonction appropriée à la mesure de réponse d'intérêt. L'habituation à court-terme peut alors être provoquée avec un poids  $w_2 = 0$ , transformant ainsi l'état  $A_{2S}$  en un état "réfractaire" qui permet d'isoler les éléments représentant  $S$  pendant un certain temps après la présentation de  $S$ . Ainsi, lorsque le stimulus est présenté à nouveau peu de temps après, ces éléments ne peuvent plus contribuer au comportement de réponse à travers  $A_{1S}$ , diminuant l'intensité de la réponse.

Dans URIBE-BAHAMONDE et al., 2019, les auteurs utilisent le SOP pour simuler les dix caractéristiques de l'habituation. Pour les caractéristiques 3 (potentialisation de l'habituation) et 10 (habituation à long terme), ils utilisent les mécanismes du modèle qui permettent l'association entre le stimulus répété et le contexte qui sert alors, dans sa globalité, de stimulus conditionnel  $SC$ .

Les travaux de GROVES et THOMPSON, 1970 sont les premiers modèles informatiques des mécanismes synaptiques de l'habituation et de la sensibilisation spécifique au site. Les réponses comportementales, conformes aux mesures biologiques, sont un mélange d'habituation et de sensibilisation spécifique au site, l'habituation étant plus prononcée pour les stimuli de faible intensité et la sensibilisation pour les stimuli de forte intensité.

Le modèle de STANLEY, 1976 est une version simplifiée du modèle de Groves et Thompson. Il permet de simuler les effets à court terme de l'habituation à partir de l'équation différentielle de premier ordre :

$$\tau \frac{dy(t)}{dt} = \alpha(y_0 - y(t)) - S(t) \quad (2.2)$$

avec  $y$  l'efficacité synaptique,  $y_0$  sa valeur initiale,  $S(t)$  la stimulation externe,  $\tau$  une constante de temps qui contrôle le taux d'habituation et  $\alpha$  qui contrôle le taux de récupération spontanée.

Cette équation a ensuite été affinée par WANG, 1993 en la paire d'équations suivantes afin d'incorporer les effets à long terme de l'habituation :

$$\tau \frac{dy(t)}{dt} = \alpha z(t)(y_0 - y(t)) - \beta y(t)S(t), \quad (2.3)$$

$$\frac{dz(t)}{dt} = \gamma z(t)(z(t) - 1)S(t) \quad (2.4)$$

Comme pour l'équation 2.2, le membre droit de l'équation 2.3 se décompose en deux termes, le premier régulant la récupération spontanée et le second l'habituation.  $\beta$  permet de paramétrer l'efficacité du stimulus présynaptique  $S(t)$ , et l'ajout du facteur  $y(t)$  transforme l'entrée directe de l'équation 2.2 en une entrée liée à l'activité. Cela permet d'atteindre une asymptote avant que  $y$  n'atteigne 0 ou une valeur négative. Le premier terme est modifié en le multipliant par  $z(t)$ , défini dans l'équation 2.4, ce qui permet la modification à long terme de la récupération spontanée. L'évolution de  $z(t)$  est plus lente que  $y(t)$  et elle permet de passer de manière non linéaire d'un  $z(t)$  élevé (récupération rapide) à un  $z(t)$  bas (récupération lente). Le paramètre  $\gamma$  impacte la pente de la courbe d'évolution de  $z(t)$ , avec une transition plus abrupte pour  $\gamma = 1$  que pour un  $\gamma$  se rapprochant de 0. L'efficacité synaptique est toujours  $y$ .



Le modèle neuro-physiologique de Hawkins (Robert D. HAWKINS, 1989) peut rendre compte d'une variété de mécanismes d'apprentissage : habituation, pseudo-conditionnement et conditionnement classique. L'auteur apporte la suggestion que les apprentissages les plus simples constituent le terreau de l'émergence d'autres formes d'apprentissages plus complexes, notamment que le conditionnement classique pourrait provenir du pseudo-conditionnement. Les résultats de ses simulations supportent cette hypothèse dans le cadre des circuits neuronaux simples.

### 2.2.2 Utilisation de l'habituation pour la détection de nouveauté

Des modèles plus opérationnels ont été proposés, la plupart utilisant un formalisme de type neuronal. Nous regroupons dans un premier temps tous les travaux qui emploient l'habituation comme mécanisme permettant la détection de nouveauté en filtrant ce qui est perçu régulièrement.

Le modèle de MARSLAND, 1999 explore une approche de modélisation de l'habituation qui consiste à incorporer une mémoire sensorielle qui s'accumule avec la présentation répétée d'un stimulus et qui module finalement l'expressivité d'une réponse motrice. Dans leur article, les auteurs considèrent à la fois un intégrateur linéaire pour moduler l'efficacité synaptique de leurs neurones modélisés et un modèle non linéaire introduit par WANG et HSU, 1990 qui peut rendre compte des effets d'habituation à long terme. Leur modèle neuronal d'habituation est appliqué à un robot mobile et l'habituation est observée dans le comportement du robot : il finit par ne plus répondre à des stimuli répétés. Le modèle montre une hiérarchie de l'habituation (l'entraînement avec un stimulus d'une intensité donnée entraîne l'habituation à des stimuli d'intensité plus faible, mais pas à des stimuli d'intensité plus élevée). Les auteurs concluent que leur modèle est intéressant pour la plausibilité biologique, mais moins pour un robot ayant d'autres tâches à effectuer du fait du coût de calcul. Ils proposent que l'habituation en robotique pourrait à la fois servir dans l'apprentissage à ignorer les stimuli "bruyants" (non nécessaire à sa tâche), mais aussi dans la détection de nouveauté, qu'ils explorent dans leurs travaux suivants (Stephen MARSLAND, NEHMZOW, SHAPIRO et al., 2000).

Dans ce second papier, leur objectif est donc d'utiliser l'habituation comme un moyen de détecter la nouveauté en apprenant ce qui est "normal" dans l'environnement du robot. L'idée est que tout stimulus perçu régulièrement est considéré comme faisant parti de la norme du contexte actuel et est par conséquent de moins en moins saillant. Un stimulus inhabituel, qu'il soit nouveau ou suffisamment rare, devient par défaut plus perceptible. Ce second modèle est basé sur leur carte auto-organisatrice *Habituating Self-Organizing Map* (HSOM, introduite dans ce même article) comme filtre de nouveauté utilisant l'habituation, et ils l'expérimentent sur un robot mobile Nomad 200 explorant des couloirs.

Stephen MARSLAND, 2009 explore comment les modèles précédents peuvent aborder des problèmes tels que la détection de la nouveauté, la détection de la récence et l'apprentissage temporel ("*learning when the temporal context of inputs is important*"). La détection de la nouveauté est possible, de même que la détection de la récence (détection d'une occurrence antérieure récente d'un événement). Cependant, l'apprentissage temporel est discuté mais n'est pas abordé expérimentalement dans cet article. La carte auto-organisatrice utilisée est *Grow When Required* (GRW, Stephen MARSLAND, SHAPIRO et NEHMZOW, 2002), une version améliorée de la HSOM. Dans certains cas, en particulier à cause du nombre fini de nœuds de la carte, la HSOM pouvait saturer et classer des stimuli nouveaux comme ayant déjà été perçus. La GRW permet de résoudre ce problème en permettant l'ajout et la suppression des arrêtes et des nœuds pendant

l'apprentissage.

L'habituation est aussi utilisée comme détection de nouveauté dans les travaux de BREAZEAL et SCASSELLATI, 1999 sur le robot Kismet et dans ceux de RUESCH et al., 2008 sur le robot iCub. Kismet est un robot conçu pour interagir socialement avec les humains. L'objectif de ces travaux est de doter le robot d'un système d'attention, basé sur des modèles d'attention humaine et de recherche visuelle, qui lui permette de se comporter suffisamment comme un humain pour maintenir une interaction sociale normale. Le rôle de l'habituation dans ces travaux est de forcer le changement de focus de l'attention visuelle depuis la cible préférentielle vers l'autre cible (soit d'une face humaine à un jouet ou *vice versa*). Pour le robot iCub, RUESCH et al., 2008 proposent un système d'attention multimodale (visuelle et auditive). Ce système utilise une *sphère égocentrique*, une représentation centrée sur le robot de l'information saillante. Pour construire cette représentation, les stimuli visuels et auditifs sont d'abord traités avec des cartes de saillance, puis ces cartes sont fusionnées. Ils utilisent l'habituation dans un mécanisme d'inhibition dynamique qui incite le robot à explorer son environnement à partir des stimuli les plus saillants vers ceux les moins saillants. L'habituation et l'inhibition sont aussi représentées *via* des sphères égocentriques. Plus le focus de l'attention visuelle reste sur une même région, plus la valeur associée à cette région dans la carte de l'habituation augmente dans  $[0; 1]$ . Cette information est ensuite utilisée dans la carte d'inhibition, qui garde ainsi une trace des régions déjà explorées, avec un paramètre de décomposition dans le temps permettant la désinhibition.

### 2.2.3 Autres travaux opérationnels

Les derniers travaux opérationnels que nous référençons n'ont pas d'objectifs communs, mais utilisent tous de l'apprentissage non-associatif dans leurs modèles.

BI et al., 2017 proposent des neurones dotés d'habituation. L'activité de ces neurones est inhibée lorsqu'ils reçoivent fréquemment un vecteur d'entrée proche de leur vecteur d'entrée moyen. Dans le contexte des réseaux neuronaux à convolution profonde, les auteurs insèrent des couches d'habituation entre les couches de produits internes de la tête du réseau. Ils rapportent une augmentation des performances de ces réseaux capable d'habituation.

Dans HONG et al., 2020, l'habituation et la sensibilisation spécifique au site sont toutes deux modélisées pour un processeur neuromorphique. Un stimulus faible et répété entraîne l'habituation de l'activité des neurones, tandis qu'un stimulus fort entraîne une sensibilisation spécifique au site de l'activité des neurones. Ils montrent que cette architecture conduit à un matériel plus robuste avec de meilleures propriétés de vieillissement.

Dans les travaux sur *Attentive and Self-Modifying cognitive architecture (ASMO)* (voir NOVIANTO, JOHNSTON et WILLIAMS, 2013), l'habituation et la sensibilisation spécifique au site utilisent essentiellement une fonction de "signification" du stimulus plutôt que son intensité, ce qui permet l'habituation aux stimuli forts et la sensibilisation spécifique au site aux stimuli faibles (effet inverse des résultats classiques en biologie). Cela permet à leur robot de rester concentré sur sa cible en favorisant les stimuli proches de la cible (jugés significatifs) par rapport à d'autres plus facilement ignorés.

Dans les travaux sur l'agent *Self-Aware and Self-Effecting (SASE)* (voir WENG, 2007), les auteurs formalisent six architectures *mentales*, de la plus simple à la plus complexe, toutes basées sur des processus de décision de Markov axés sur l'observation. À partir des architectures de

type 4, leurs modèles conceptuels sont censés permettre l'habituation et la sensibilisation spécifique au site, mais cette architecture de type 4 est également compatible avec le conditionnement classique, le conditionnement instrumental (conditionnement opérant), la planification et le raisonnement. Par ailleurs, seules les architectures les plus simples (jusqu'au type 2) sont expérimentées, l'habituation et la sensibilisation nécessitant au moins une architecture de type 4.

Dans le *Autonomous Robot Based on Inspirations from Biology* (ARBIB, voir DAMPER, FRENCH et SCUTT, 2000), l'habituation et le pseudo-conditionnement sont modélisés par un réseau de neurones à impulsions, ainsi que le conditionnement classique. Les neurones codent les réflexes, et sont implémentés sur un robot Khepera pour observer les comportements résultants. Conscients de la complexité déconcertante des modèles trop fidèles à la modélisation physiologique, Damper et al. se sont appuyés sur des réseaux de neurones à impulsions pour abstraire les mécanismes physiologiques, mais il s'agit toujours d'un modèle assez complexe. Il n'est pas clair non plus dans quelle mesure leur modèle peut exploiter l'apprentissage non-associatif sans s'appuyer sur les mécanismes nécessaires au conditionnement classique.

## 2.3 Bilan de l'état de l'art

Les apprentissages non-associatifs nous semblent être un bon point de départ pour permettre à un agent de progressivement agrandir son répertoire comportemental de manière non supervisée et non aléatoire.

Plusieurs travaux de modélisations de ces apprentissages existent déjà. Certains cherchent à simuler des résultats biologiques pour appuyer des hypothèses sur les mécanismes dans le vivant, d'autres utilisent les apprentissages non-associatifs comme inspiration pour tirer profit de leurs caractéristiques (système de détection de nouveauté ou système d'attention par exemple).

Pour cette thèse, nous avons préféré utiliser sur le modèle *Iterant Deformable Sensorimotor Medium* (IDSM, issu des travaux de M. D. EGBERT et BARANDIARAN, 2014 et présenté formellement dans le chapitre suivant). Ses caractéristiques qui ont motivés ce choix sont qu'il propose un cadre déjà existant pour représenter les comportements à un niveau de description mésoscopique. Ce niveau est intermédiaire entre un niveau de description fin comme les architectures neuronales et les descriptions d'actions ou mouvements macroscopiques. Cela présente l'avantage d'être plus interprétable que les réseaux de neurones, dans le sens où les analyses peuvent être plus simples ou plus intuitives. De plus, les comportements sont représentés de manière continue dans un espace sensorimoteur plutôt que d'être basés sur un ensemble prédéfini d'actions discrétisées, ce qui offre plus de possibilités comportementales à l'agent. Enfin, nous rappelons que nous cherchons à nous inspirer des apprentissages non-associatifs en biologie sans essayer de retrouver précisément des résultats ou observations issues d'expériences sur le vivant. Par exemple, pour l'habituation, ce qui nous intéresse est que l'agent soit capable de diminuer sa réponse à un stimulus lorsque le stimulus est répété plutôt que de modéliser fidèlement les mécanismes utilisés par un organisme unicellulaire comme le blob ou ceux utilisés par un neurone.

Le prochain chapitre décrit formellement l'IDSM, puis présente des expériences reproduites de M. D. EGBERT et BARANDIARAN, 2014 ainsi que des expériences originales que nous avons faites en amont de nos principales contributions avec les apprentissages non-associatifs.



# 3

## IDSMS

Les travaux de cette thèse utilisent comme substrat numérique l'*Iterant Deformable Sensorimotor Medium* (IDSMS, présenté dans M. D. EGBERT et BARANDIARAN, 2014). C'est un modèle dynamique abstrait encodant des comportements répétitifs à échelle mésoscopique, intermédiaire entre microscopique (modèles basés sur les neurones par exemple) et macroscopique (modèles basés sur des comportements discrétisés en actions par exemple). Ses auteurs le présentent comme un outil conceptuel sans plausibilité biologique pour réfléchir sur les habitudes en tant que tendances à ré-exprimer des trajectoires sensorimotrices déjà rencontrées. Les éléments centraux sur lesquels repose l'IDSMS sont des *nœuds sensorimoteurs* qui conservent l'information de ce que l'agent a déjà expérimenté dans l'espace sensorimoteur SM et qui influencent son état moteur pour qu'il tende à répéter ces expériences sensorimotrices.

La figure 3.1 schématise l'interaction entre l'espace sensorimoteur SM où fonctionne l'IDSMS et l'environnement (physique ou simulé) où se trouve l'agent (physique ou simulé). L'état sensorimoteur de l'agent  $\mathbf{x}_{SM}$  est calculé à partir de la perception par l'agent de l'état de l'environnement  $S_E$ . L'état  $\mathbf{x}_{SM}$  est constitué de la composante motrice  $[\mathbf{x}_{SM}]^\mu$  et de la composante sensorielle  $[\mathbf{x}_{SM}]^\sigma$  qui contiennent respectivement les informations normalisées des moteurs et celles des senseurs de l'agent. Dans l'espace sensorimoteur SM, l'IDSMS utilise à la fois les *nœuds* qui

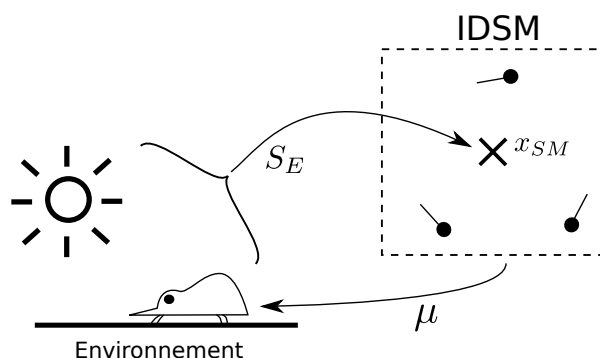


FIGURE 3.1 – L'IDSMS est basé sur l'espace sensorimoteur normalisé de l'agent. À partir de la position sensorimotrice actuelle de l'agent  $\mathbf{x}_{SM}$  dans cet espace et des différents nœuds de l'IDSMS, la commande motrice  $\mu$  est calculée. Cette commande contrôle le comportement de l'agent dans son environnement.  $S_E$  est l'état complet de l'environnement.

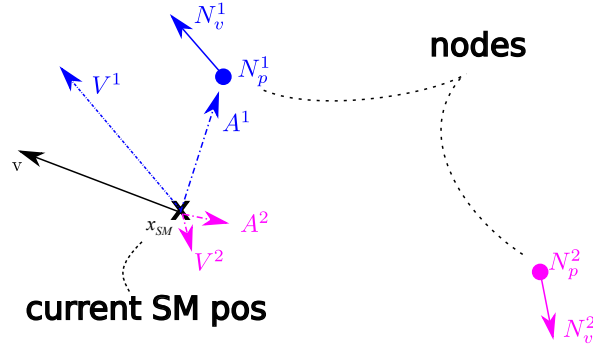


FIGURE 3.2 – Un IDSM est construit à partir d’un ensemble de nœuds  $N^i$  qui influencent la dynamique de l’agent dans l’espace sensorimoteur SM à travers des composantes vitesse  $\mathbf{V}^i$  et attraction  $\mathbf{A}^i$ .

le composant (chacun contenant une information de dynamique locale) et  $\mathbf{x}_{SM}$  pour calculer la commande motrice  $\boldsymbol{\mu}$ . L’application de cette commande  $\boldsymbol{\mu}$  modifie la composante motrice  $[\mathbf{x}_{SM}]^\mu$ , qui est dénormalisée pour mettre à jour les moteurs, ce qui impacte l’agent dans son environnement  $\mathbf{S}_E$  et éventuellement l’état de l’environnement  $S_E$ .

La figure 3.2 schématise l’influence de chaque nœud exercée sur la vitesse de l’agent dans l’espace sensorimoteur SM. Cette influence se décline en deux composantes qui dépendent chacune de l’ensemble des éléments du nœud (définis formellement en 3.1) : une composante d’attraction  $\mathbf{A}^i$ , cherchant à rapprocher l’agent de la position du nœud  $\mathbf{N}_p$ , et une composante de vitesse  $\mathbf{V}^i$ , cherchant à aligner la vitesse de l’agent avec la vitesse stockée dans le vecteur  $\mathbf{N}_v$  du nœud. L’amplitude des composantes  $\mathbf{A}^i$  et  $\mathbf{V}^i$  est fonction de la distance entre les positions de l’agent  $\mathbf{x}_{SM}$  et du nœud  $\mathbf{N}_p$  (plus l’agent est proche plus le nœud est influent) et du poids du nœud  $N_w$ . Les nœuds vont avoir ainsi tendance à attirer l’agent vers leurs positions et à biaiser la trajectoire de l’agent pour qu’elle s’aligne avec leurs vitesses.

### 3.1 Modèle

L’espace sensorimoteur SM utilisé par l’IDSM a autant de dimensions que l’agent a de moteurs et de senseurs : on note  $K = \dim(SM)$  le nombre de dimensions. Chacune de ces dimensions motrices ou sensorielles est normalisée dans  $[0, 1]$ .

Concernant les notations, nous représentons les vecteurs avec des caractères en gras (par exemple  $\mathbf{v}$ ) ou entre crochets (par exemple  $[f, 0, 0]$ ).  $[\mathbf{v}]^\mu$  est la projection de  $\mathbf{v}$  sur les dimensions du moteur, et  $[\mathbf{v}]^\sigma$  est la projection sur les dimensions du senseur. En outre,  $\mathbf{1}$  est le vecteur dont toutes les composantes valent 1 et  $\langle \dots, \dots \rangle$  désigne le produit scalaire.

Un nœud  $N$  de l’IDSM est formellement défini par un quadruplet :

$$N = \langle \mathbf{p}, \mathbf{v}, w, t \rangle \quad (3.1)$$

avec

- $\mathbf{p}$  un vecteur qui indique la position du nœud  $N$  dans l’espace sensorimoteur SM,
- $\mathbf{v}$  un vecteur qui indique la vitesse souhaitée dans l’espace sensorimoteur SM à la position  $\mathbf{p}$ ,

- $w$  un scalaire indiquant le poids du nœud, qui permet de pondérer son influence relativement aux autres nœuds lors du calcul de la dynamique de l'agent dans l'espace sensorimoteur SM,
- $t$  un scalaire indiquant la date de création du nœud.

La date de création du nœud  $t$  n'est pas indiquée dans l'article M. D. EGBERT et BARANDIARAN, 2014, mais nous l'avons ajoutée pour clarifier le calcul de l'activation de l'influence du nœud, expliqué par la suite. Les différentes composantes du nœud  $N$  sont désignées avec une notation en indice :  $\mathbf{N}_p$ ,  $\mathbf{N}_v$ ,  $N_w$ ,  $N_t$ .

Lors du parcours de l'agent dans l'espace sensorimoteur, un nouveau nœud est créé lorsque la densité des nœuds  $\phi_{\mathcal{N}}$  à la position sensorimotrice actuelle de l'agent  $\mathbf{x}_{SM}$  est inférieure à un seuil défini par le scalaire  $k_t = 1$ . Nous avons rajouté un paramètre  $\Delta t_{nodes}$  permettant de dissocier le temps des événements de création de nœuds du temps de simulation. Ce paramètre est l'écart de temps entre chaque tentative de création de nœud, donc entre chaque vérification que la densité des nœuds  $\phi_{\mathcal{N}}$  est inférieure ou non au seuil  $k_t$ . Par défaut, nous fixons  $\Delta t_{nodes} = \Delta t = 0.1$ , donc la création de nœud se fait dès que possible.

La densité des nœuds est donnée par les équations 3.2-3.4 :

$$\phi_{\mathcal{N}}(\mathbf{x}_{SM}) = \sum_{N \in \mathcal{N}} \omega(N_w) \cdot d(\mathbf{N}_p, \mathbf{x}_{SM}) \quad (3.2)$$

$$\omega(N_w) = \frac{2}{1 + \exp(-k_w N_w)} \quad (3.3)$$

$$d(\mathbf{N}_p, \mathbf{x}_{SM}) = \frac{2}{1 + \exp(k_d \|\mathbf{N}_p - \mathbf{x}_{SM}\|^2)} \quad (3.4)$$

avec  $\mathcal{N}$  l'ensemble de tous les nœuds,  $k_d = 1000$  et  $k_w = 0.0025$  par défaut. Les équations du facteur de distance (équation 3.4) et du facteur de poids (équation 3.3) sont illustrées dans la figure 3.3. La figure 3.4 donne des exemples de la densité  $\phi_{\mathcal{N}}$  pour un IDSM avec une unique dimension.

Lors de sa création, les composantes du nouveau nœud  $N$  sont :

- $\mathbf{N}_p = \mathbf{x}_{SM}$  ;
- $\mathbf{N}_v$  est une estimation de la vitesse actuelle de l'agent par les différences finies ;
- $N_w = 0$  ;
- $N_t = t$  avec  $t$  le temps de simulation à la création du nœud.

Le poids  $N_w$  du nœud  $N$  évolue en fonction du temps directement après sa création suivant l'équation différentielle 3.5. Le poids diminue de manière indépendante de l'agent et constamment dans le temps. Il augmente en fonction de la distance entre le nœud  $N$  et la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$ , de manière plus importante lorsque cette distance est petite. Le poids augmente si  $d(\mathbf{N}_p, \mathbf{x}_{SM}) \geq 0.1$ , donc si la distance  $\|\mathbf{N}_p - \mathbf{x}_{SM}\| \leq 0.054$  (voir la figure 3.3 pour la courbe de  $d$ ) avec les paramètres par défaut pour l'équation 3.4. Comme nous le verrons plus tard, au delà de leur impact dans le calcul de densité, ces poids modulent également la contribution d'un nœud à la commande motrice de l'agent. Plus le poids est fort, plus le nœud influencera la trajectoire de l'agent (équations 3.8,3.9). La distance va également impacter l'importance de l'influence d'un nœud mais le poids est un accumulateur qui reflète la fréquence à laquelle l'agent passe à proximité de ce nœud. Ainsi, plus souvent l'agent passe proche d'un nœud, plus ce nœud

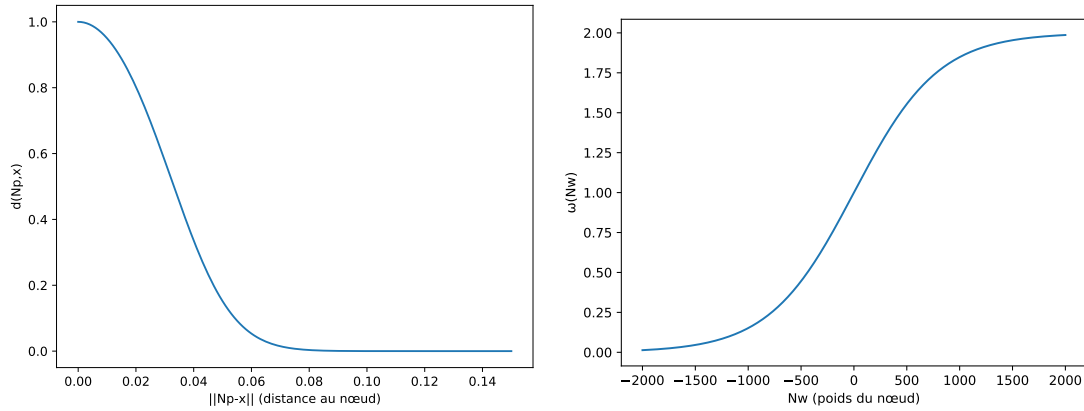


FIGURE 3.3 – Fonctions non linéaires utilisées pour calculer la densité de nœuds d’un état sensorimoteur, et pour mettre à l’échelle l’influence des nœuds en fonction de leur proximité avec l’état sensorimoteur actuel (à gauche), selon l’équation 3.4, et de leur poids (à droite), selon l’équation 3.3.

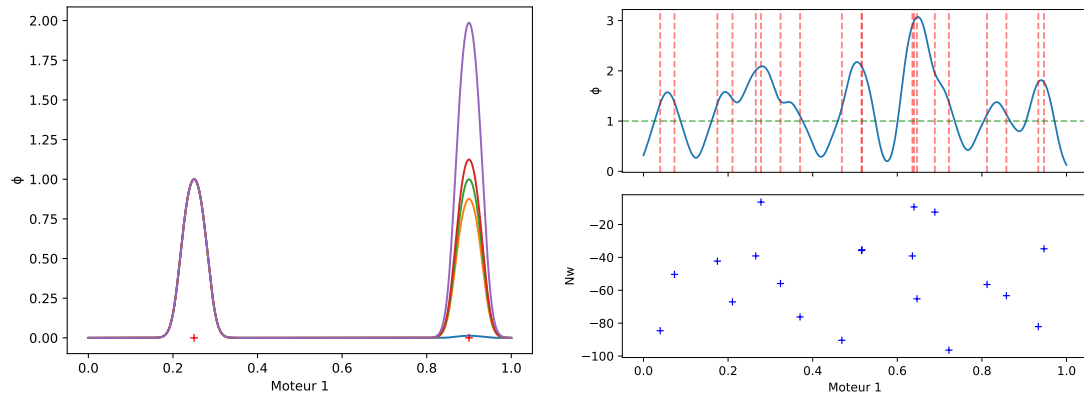


FIGURE 3.4 – Affichage de la densité  $\phi_{\mathcal{N}}$  pour un IDSM en une seule dimension. Le graphique de gauche montre l’impact de la variation du poids du nœud  $N^2$  en  $\mathbf{N}_p^2 = [0.9]$ . Son poids  $N_w^2$  vaut  $-2000$  (en bleu),  $-100$  (orange),  $0$  (vert),  $100$  (rouge) et  $2000$  (violet). Le poids du premier nœud  $N^1$  en  $\mathbf{N}_p^1 = [0.25]$  reste à  $0$ . Le second graphique donne la densité pour 20 nœuds à positions et poids aléatoires (position dans  $[0, 1]$ , poids dans  $[-100, 0]$ ) pour donner un exemple *théorique* d’un paysage de densité  $\phi_{\mathcal{N}}$ . Les lignes verticales rouges représentent la position des nœuds, la ligne horizontale verte indique le seuil par défaut  $k_t = 1$  en dessous duquel la création de nœud serait possible. Dans cet exemple aléatoire, la création de nœud n’est pas possible dans la majorité de l’espace.



est influent.

$$\frac{dN_w}{dt} = -1 + 10.d(\mathbf{N}_p, \mathbf{x}_{SM}) \quad (3.5)$$

Les équation différentielles utilisées pour la dynamique du modèle sont intégrées avec la méthode d'Euler avec  $\Delta t = 0.1$ .

À chaque pas de temps, un sous-ensemble de nœuds, les nœuds *actifs*, influencent la commande motrice de l'agent. Est considéré comme actif un nœud ayant été créé depuis un certain temps  $\tau$ . Ce sous-ensemble de nœuds actifs au temps  $t$  est noté  $\mathcal{A}(t)$ , ou  $\mathcal{A}$  lorsque  $t$  est implicite. La période d'inactivation d'un nœud juste après sa création sert à tempérer son influence, sinon sa proximité avec l'agent aurait tendance à supplanter l'influence des autres nœuds.

$$\mathcal{A}(t) = \{N \in \mathcal{N} | N_t > \tau\} \quad (3.6)$$

La fonction de densité des nœuds actifs est notée  $\phi_{\mathcal{A}}$  et est définie par l'équation :

$$\phi_{\mathcal{A}}(\mathbf{x}_{SM}) = \sum_{N \in \mathcal{A}} \omega(N_w).d(\mathbf{N}_p, \mathbf{x}_{SM}) \quad (3.7)$$

La différence entre les fonctions de densité  $\phi_{\mathcal{N}}$  et  $\phi_{\mathcal{A}}$  est donc quel ensemble de nœuds est utilisé lors de leur calcul respectif. La création d'un nœud dépend de la densité  $\phi_{\mathcal{N}}$  calculée à partir de l'ensemble des nœuds  $\mathcal{N}$  alors que la commande motrice que l'agent applique dépend de la densité  $\phi_{\mathcal{A}}$  calculée uniquement à partir des nœuds actifs  $\phi_{\mathcal{A}}$ .

La dynamique de l'influence des nœuds pour la position sensorimotrice actuelle de l'agent  $\mathbf{x}_{SM}$  est composée de deux facteurs : un vecteur de vitesse  $\mathbf{V}$  et un vecteur d'attraction  $\mathbf{A}$ .  $\mathbf{V}$  est l'impact qu'ont tous les  $\mathbf{N}_v$  des nœuds actifs sur l'agent, relativement à leurs poids et à la position sensorimotrice actuelle de l'agent  $\mathbf{x}_{SM}$ . C'est ce facteur de vitesse qui tend à forcer l'agent à reproduire des trajectoires qu'il a déjà suivies.  $\mathbf{A}$  est l'impact de la position des nœuds actifs sur l'agent, aussi de manière relative à leurs poids et à la position sensorimotrice actuelle de l'agent  $\mathbf{x}_{SM}$ . Ce terme d'attraction modifie la commande motrice pour rapprocher la position sensorimotrice actuelle de l'agent  $\mathbf{x}_{SM}$  de la zone de l'espace sensorimoteur la plus dense en nœuds. Le vecteur  $\mathbf{\Gamma}$  propre à chaque nœud  $N$  permet d'éviter que le facteur d'attraction résultant aille à l'encontre de la direction proposée par le vecteur vitesse  $\mathbf{N}_v$  en supprimant du vecteur d'attraction toute composante colinéaire avec  $\mathbf{N}_v$ .

$$\mathbf{V}(\mathbf{x}_{SM}) = \sum_{\mathcal{N}} \omega(N_w).d(\mathbf{N}_p, \mathbf{x}_{SM}).[\mathbf{N}_v]^\mu \quad (3.8)$$

$$\mathbf{A}(\mathbf{x}_{SM}) = \sum_{\mathcal{N}} \omega(N_w).d(\mathbf{N}_p, \mathbf{x}_{SM}).[\mathbf{\Gamma}(\mathbf{N}_p - \mathbf{x}_{SM}, \mathbf{N}_v)]^\mu \quad (3.9)$$

$$\mathbf{\Gamma}(\mathbf{x}, \mathbf{N}_v) = \begin{cases} \mathbf{x} - \left\langle \mathbf{x}, \frac{\mathbf{N}_v}{\|\mathbf{N}_v\|} \right\rangle \frac{\mathbf{N}_v}{\|\mathbf{N}_v\|} & \text{si } \mathbf{N}_v > \epsilon \\ \mathbf{x} & \text{sinon} \end{cases} \quad (3.10)$$

La notation  $[\mathbf{x}]^\mu$  indique la projection du vecteur  $\mathbf{x}$  dans la dimension motrice de l'espace sensorimoteur SM. La fonction  $\mathbf{\Gamma}(\mathbf{x}, \mathbf{y})$  conserve uniquement la composante de  $\mathbf{x}$  perpendiculaire à  $\mathbf{y}$ .

La commande motrice  $\boldsymbol{\mu}$  de l'agent permet de modifier les moteurs de l'agent à travers la composante motrice  $[\mathbf{x}_{SM}]^\mu$  (état des moteurs normalisés dans l'espace sensorimoteur), calculée en bornant la valeur de  $\boldsymbol{\mu}$  dans  $[0, 1]$ . Cette commande motrice évolue sous l'influence des facteurs  $\mathbf{V}$  et  $\mathbf{A}$  :

$$\dot{\boldsymbol{\mu}}(t) = \frac{d\boldsymbol{\mu}}{dt} = \begin{cases} \frac{\mathbf{V}(\mathbf{x}_{SM}) + \mathbf{A}(\mathbf{x}_{SM})}{\phi_{\mathcal{A}}(\mathbf{x}_{SM})} & \text{si } \phi_{\mathcal{A}}(\mathbf{x}_{SM}) > \epsilon \\ 0 & \text{sinon} \end{cases} \quad (3.11)$$

## 3.2 Exemples d'IDSM simples

$\Delta t$	$k_t$	$k_d$	$k_\omega$	$\tau$	$\boldsymbol{\mu}$
0.1	1	1000	0.0025	10	$\in [-\infty; +\infty]$

TABLE 3.1 – Valeurs des paramètres et options utilisées pour les exemples d'IDSM simples (section 3.2)

Pour donner une meilleure intuition de ce que permet un IDSM, nous reproduisons des exemples utilisés dans l'article de M. D. EGBERT et BARANDIARAN, 2014. L'intégralité de cette réplcation est publiée dans notre article auprès du journal ReScience (GILLARD, FIX et DUTECH, 2022).

Dans chacun de ces exemples, le comportement initial est provoqué par une marche forcée de l'agent (la commande motrice est prédéfinie et indépendante de l'IDSM), fournissant ainsi un ensemble de nœuds qui devraient inciter au même comportement. Ensuite, l'IDSM est laissé libre et la dynamique des nœuds dans l'espace sensorimoteur devrait alors forcer l'agent à reprendre le comportement attendu.

### 3.2.1 Cercle

Le premier exemple utilise un agent doté de deux moteurs et d'aucun senseur (voir figure 3.5). Le comportement provoqué est qu'il tourne en rond. Pour cela, les équations utilisées pendant l'entraînement pour la marche forcée sont :

$$m_1(t) = 0.75 \cos\left(\frac{2\pi t}{10}\right) \quad (3.12)$$

$$m_2(t) = 0.75 \sin\left(\frac{2\pi t}{10}\right) \quad (3.13)$$

L'entraînement dure pendant 20 unités de temps pendant lesquelles des nœuds sont créés sur la trajectoire sensorimotrice de l'agent, selon la densité  $\phi_{\mathcal{N}}$  locale. Tous les nœuds d'entraînement sont alors activés et l'agent est relancé sous contrôle de l'IDSM et à partir d'une position motrice placée aléatoirement dans l'espace moteur.

Dans la figure 3.5, le premier graphique de 3.5a illustre la trajectoire suivie par l'agent pendant l'entraînement ainsi que la dynamique résultante des nœuds à la fin de celui-ci. Chaque flèche est la variation de la commande motrice  $\dot{\boldsymbol{\mu}}(t)$  calculée par l'IDSM si l'agent se trouvait à cette

position. Sur le second graphique, les différents nœuds créés pendant l'entraînement sont représentés. Pour un nœud  $N$ ,  $\mathbf{N}_p$  est la base de la flèche,  $\mathbf{N}_v$  est la taille et l'orientation de la flèche, et le poids du nœud relatif aux autres est représenté par l'intensité de la couleur. Les graphiques de 3.5b montrent les trajectoires suivies par l'agent après relocalisation dans l'espace moteur en  $(0.13, 0.17)$ . Le graphique de gauche montre la dynamique obtenue en fin d'expérience avec la trajectoire d'entraînement en bleu et celle pendant l'IDSM en rouge (la trajectoire bleue est très peu visible du fait que la trajectoire rouge repasse presque parfaitement dessus). Le graphique de droite montre les trajectoires de l'agent résultantes dans son environnement. Les graphiques de 3.5c sont similaires mais avec une relocalisation en  $(0.51, 0.5)$ , proche du centre dans l'espace moteur. Dans les deux cas, l'agent dirigé par l'IDSM retrouve une trajectoire identique à celle d'entraînement : il tourne en rond. L'IDSM est donc capable de mémoriser une trajectoire sensorimotrice lorsque l'agent en fait l'expérience et de lui faire suivre à nouveau une trajectoire sensorimotrice déjà acquise.

### 3.2.2 Agent avec deux moteurs et deux senseurs

Dans les exemples suivants, l'agent possède deux moteurs et deux senseurs, simulant un robot avec deux roues et deux senseurs de lumière (représenté sur la figure 3.6). L'agent est un cercle de rayon  $r = 0.25$  à la position  $(x, y)$ , avec les senseurs disposés sur le cercle à un angle  $\pm\beta$  depuis son orientation  $\alpha$ . Avec  $\mathbf{m} = (m_l, m_r)$  la vitesse de ses roues gauche et droite, la dynamique de l'agent dépend des équations différentielles classiques d'un robot à roues différentielles :

$$\dot{x} = \cos(\alpha)(m_l + m_r), \quad (3.14)$$

$$\dot{y} = \sin(\alpha)(m_l + m_r), \quad (3.15)$$

$$\dot{\alpha} = \frac{m_r - m_l}{2r} = 2(m_r - m_l). \quad (3.16)$$

L'agent se déplace dans une arène torique de largeur 4 :  $[-2, 2]^2$ . L'activité sensorielle dépend de la distance  $D$  entre le senseur et la lampe située en  $(x = 0, y = 0)$ , ainsi que de l'angle entre la direction du senseur (le vecteur  $\mathbf{b}$ , voir figure 3.6) et la direction de la lumière vue depuis le senseur (le vecteur  $\mathbf{c}$ , voir figure 3.6). Cette activité sensorielle est calculée suivant :

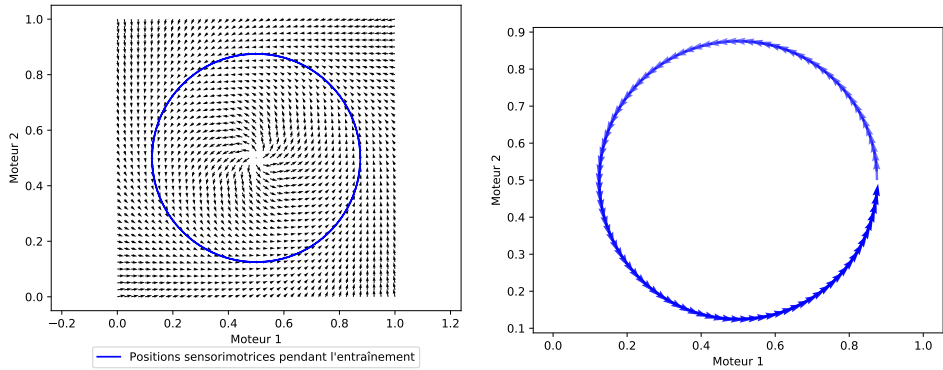
$$s = \left[ \frac{\langle \mathbf{b}, \mathbf{c} \rangle}{\|\mathbf{b}\| \|\mathbf{c}\|} \right]^+ \frac{1}{1 + D^2} \quad (3.17)$$

où  $[x]^+$  vaut soit 0 soit la valeur positive de  $x$  et  $\mathbf{b} = (\cos(\alpha + \beta), \sin(\alpha + \beta))$ .

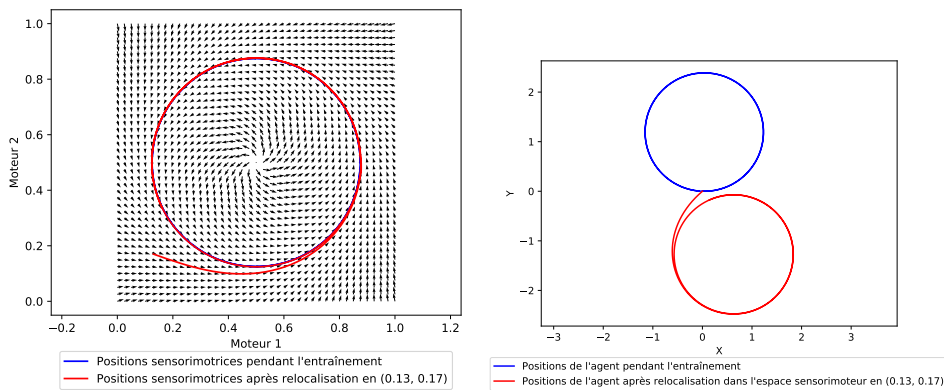
L'espace sensorimoteur de l'agent est composé de quatre dimensions, deux motrices et deux sensorielles, la position sensorimotrice de l'agent est :

$$\mathbf{x}_{SM} = F \begin{pmatrix} m_l \\ m_r \\ s_l \\ s_r \end{pmatrix} \quad (3.18)$$

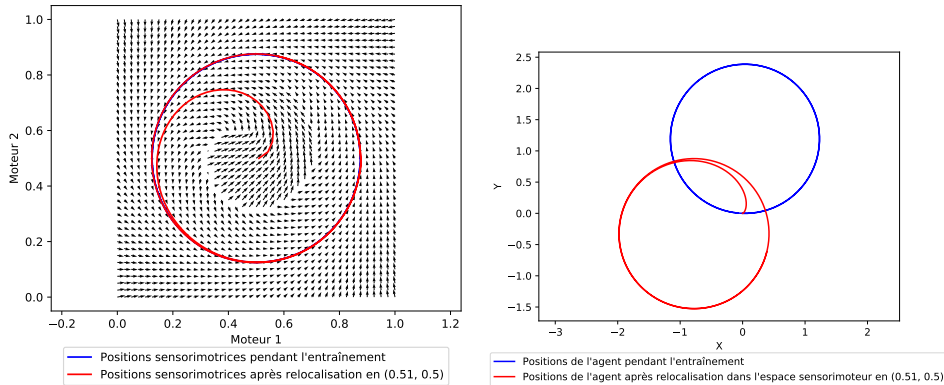
où  $f$  est la fonction transformant les valeurs des moteurs et senseurs du monde physique en position sensorimotrice de l'agent  $\mathbf{x}_{SM}$  dans l'espace sensorimoteur. Dans ces expériences, cette



(a) Dynamique et position des nœuds après l'entraînement



(b) Trajectoires de l'agent en partant de (0.13, 0.17)



(c) Trajectoires de l'agent en partant de (0.51, 0.5)

FIGURE 3.5 – Résultats d'un IDSM permettant à un agent doté de deux moteurs de tourner en rond. Le premier graphique de 3.5a montre la dynamique de l'espace moteur en fin d'entraînement et la trajectoire que l'agent a suivi. Le second graphique indique les différents nœuds créés pendant l'entraînement, la base de la flèche étant le  $\mathbf{N}_p$ , la flèche le  $\mathbf{N}_v$  et l'intensité de la couleur le poids du nœud relatif aux autres. Le premier graphique de 3.5b (resp. 3.5c) montre la dynamique en fin d'expérience lorsque la composante motrice  $[\mathbf{x}_{SM}]^\mu$  de l'agent a été repositionné en (0.13, 0.17) (resp. (0.51, 0.5)) avant de laisser le contrôle à l'IDSM, avec les trajectoires de l'agent dans l'espace moteur. Les seconds graphiques de 3.5b et 3.5c montrent les trajectoires de l'agent dans l'environnement pendant l'entraînement et suite au repositionnement dans l'espace moteur. Dans les deux cas, l'agent reprend rapidement une trajectoire identique à celle d'entraînement.

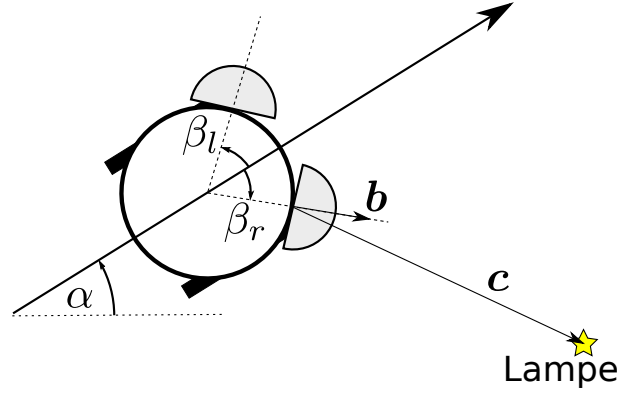


FIGURE 3.6 – Robot avec deux roues et deux senseurs de lumière.

fonction est définie par :

$$F\left(\begin{pmatrix} m_l \\ m_r \\ s_l \\ s_r \end{pmatrix}\right) = \begin{pmatrix} \frac{m_l+1}{2} \\ \frac{m_r+1}{2} \\ s_l \\ s_r \end{pmatrix} \quad (3.19)$$

Comme dans l'exemple précédent, l'agent suit d'abord une phase d'entraînement en marche forcée où ses moteurs sont contrôlés pour lui imposer un comportement initial. La commande motrice  $\chi \in [-1, 1]$  change la valeur des moteurs selon :

$$\frac{d\mathbf{m}}{dt} = (\chi - \mathbf{m}). \quad (3.20)$$

Cette commande motrice va alors impacter la position et orientation de l'agent selon les équations du mouvement 3.14-3.16.

Trois comportements différents sont illustrés : phototaxie (l'agent est attiré par la lumière), phototaxie sinusoïdale (attiré par la lumière mais avec une perturbation sinusoïdale) et photophobie (l'agent fuit la lumière tout en plaçant la lampe derrière lui).

L'entraînement dure 100 unités de temps et l'agent est repositionné aléatoirement toutes les 25 unités de temps pendant toute la durée de l'expérience. À chaque repositionnement, son état moteur est lui aussi tiré aléatoirement selon une loi uniforme. Une fois l'entraînement terminé (fin de la marche forcée), c'est l'IDSM qui pilote l'agent. La composante motrice de l'état sensorimoteur (toujours limitée dans  $[0, 1]$ ) de l'agent évolue alors selon les équations 3.11 (cette composante motrice est la valeur de  $\mu$  bornée dans  $[0, 1]$ ). Cette composante est ensuite transformée dans le monde physique en appliquant la transformation inverse  $F^{-1}$ . L'expérience est arrêtée après 500 unités de temps, soit un total de 20 trajectoires.

**Phototaxie** Pour simuler la phototaxie pendant l'entraînement, la commande motrice forcée  $\chi$  suit les équations suivantes :

$$\chi_l = 1 - 1.5\sqrt{s_l} \quad (3.21)$$

$$\chi_r = 1 - 1.5\sqrt{s_r} \quad (3.22)$$

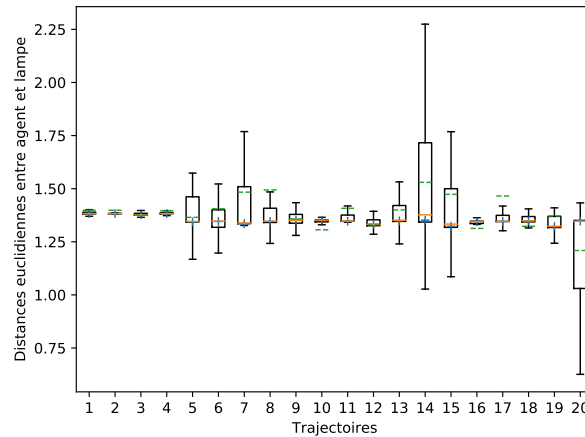
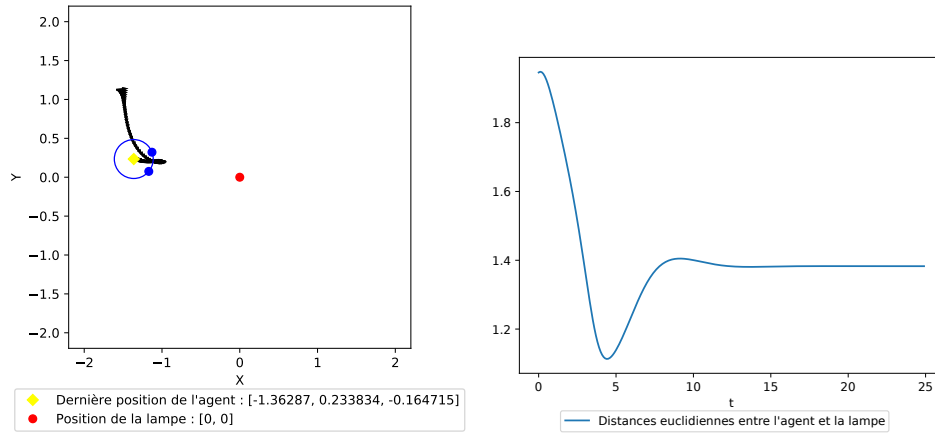


FIGURE 3.7 – Boîtes à moustache pour comparer les différences de distances entre l’agent et la lampe entre chaque trajectoire après relocalisation de l’agent pour le comportement phototaxie. Pour chaque trajectoire, la boîte est calculée à partir des 250 distances, une par pas de temps. Pour chaque boîte, le trait orange plein indique la médiane, le trait vert pointillé indique la moyenne, la croix bleue est la distance depuis la dernière position de l’agent. Les quatre premières trajectoires sont celles d’entraînement avec marche forcée. Pour toutes les trajectoires, l’agent se stabilise entre 1.3 et 1.4 unité de distance de la lampe.

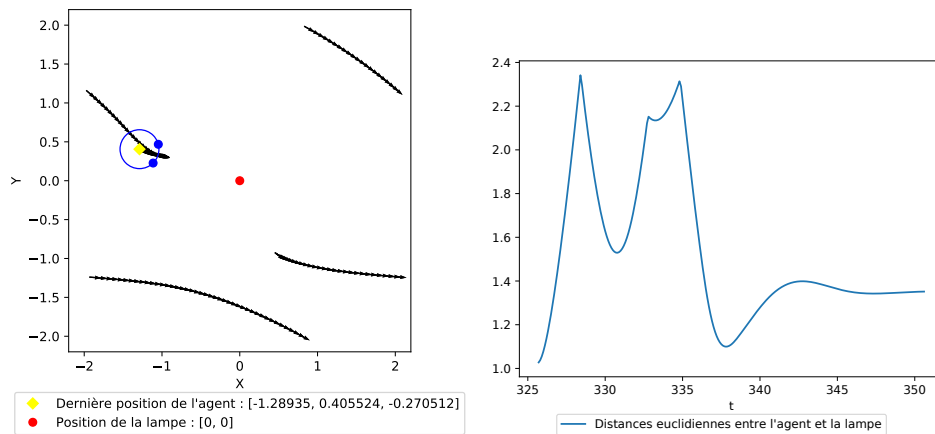
Le comportement résultant est que l’agent cherche à aller vers la lumière, et se stabilise en état de quasi-arrêt face à la lampe.

La distribution des distances de l’agent au centre de l’arène (position de la lampe) est représentée par les boîtes à moustaches de la figure 3.7. Cela donne un aperçu de l’ensemble des trajectoires effectuées par l’agent pendant l’expérience, en utilisant comme métrique la distance entre l’agent et la lampe. Pour chaque trajectoire, la boîte est calculée à partir des 250 distances qui la composent, une par pas de temps (soit  $\frac{25}{\Delta t}$ ). Pour chaque boîte, le trait orange plein indique la médiane, le trait vert pointillé indique la moyenne, la croix bleue est la distance depuis la dernière position de l’agent. La boîte contient 50% des valeurs, entre le premier et le troisième quartile. Les moustaches indiquent les valeurs *minimum* et *maximum* de la trajectoire en dehors des valeurs aberrantes (*outliers*), ces dernières ne sont pas représentées. Pour toutes les trajectoires, l’agent se stabilise entre 1.3 et 1.4 unité de distance de la lampe. Les quatre premières trajectoires sont celles d’entraînement avec marche forcée. La figure 3.8 montre quelques exemples de trajectoires effectuées par l’agent. La figure 3.8a est la première trajectoire lors de l’entraînement, c’est la marche forcée et donc aussi le comportement attendu : l’agent s’oriente face à la lampe et s’arrête à une certaine distance d’elle (entre 1.3 et 1.4). Les deux autres trajectoires (voir figures 3.8b et 3.8c) sont celles avec les boîtes les plus large sur la figure 3.7 (trajectoires 14 et 20), donc avec des distances plus dispersées. Nous pouvons voir que dans les deux cas l’agent arrive bien à se stabiliser face à la lampe. Les pics abruptes sur le graphique des distances de la figure 3.8b sont dûs au passage d’un côté à l’autre de l’arène torique.

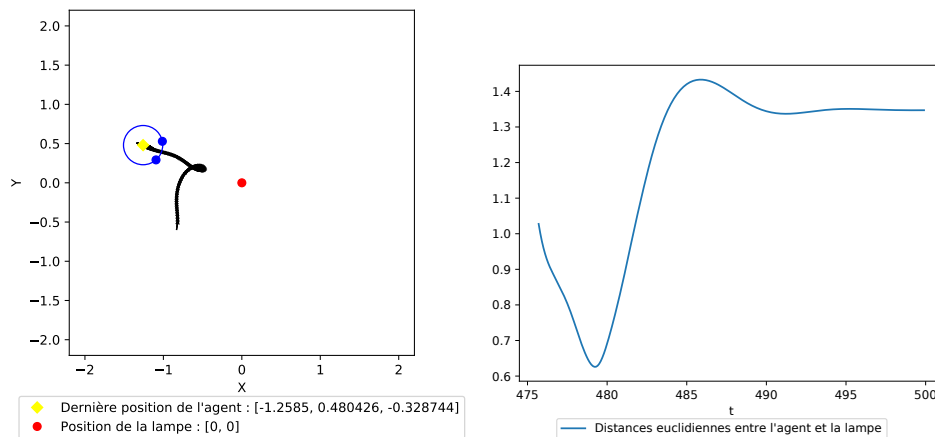
**Photophobie** Pour simuler la photophobie pendant l’entraînement, la commande motrice forcée  $\chi$  suit les équations suivantes (connections entre senseurs et moteurs inverses de celles du



(a) Première trajectoire de l'agent pendant l'entraînement.



(b) Trajectoire 14 de l'agent.



(c) Trajectoire 20 de l'agent.

FIGURE 3.8 – Quelques exemples de trajectoires de l'expérience sur le comportement phototaxie. Dans tous les cas, que ce soit pendant l'entraînement ou avec l'IDSM, l'agent arrive à se stabiliser face à la lampe à une distance entre 1.3 et 1.4 d'elle.

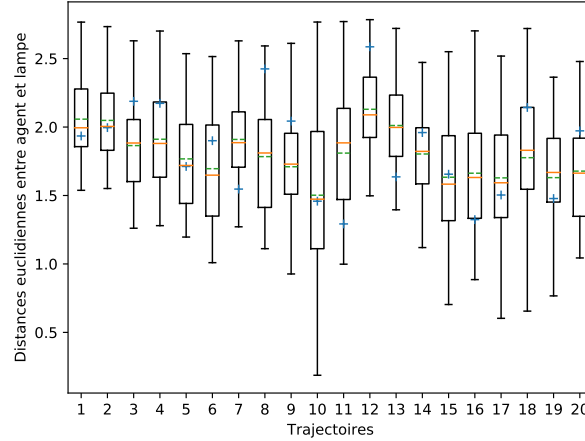


FIGURE 3.9 – Boîtes à moustache pour comparer les différences de distances entre l’agent et la lampe entre chaque trajectoire après relocalisation de l’agent pour le comportement phototaxie. Pour chaque trajectoire, la boîte est calculée à partir des 250 distances, une par pas de temps. Pour chaque boîte, le trait orange plein indique la médiane, le trait vert pointillé indique la moyenne, la croix bleu est la distance depuis la dernière position de l’agent. Les quatre premières trajectoires sont celles d’entraînement avec marche forcée. Contrairement aux résultats du comportement phototaxie, les distances sont très variées à la fois pour une même trajectoire mais aussi entre trajectoires.

comportement phototaxie) :

$$\chi_l = 1 - 1.5\sqrt{s_r} \quad (3.23)$$

$$\chi_r = 1 - 1.5\sqrt{s_l} \quad (3.24)$$

L’agent avance tout en cherchant à fuir la lampe.

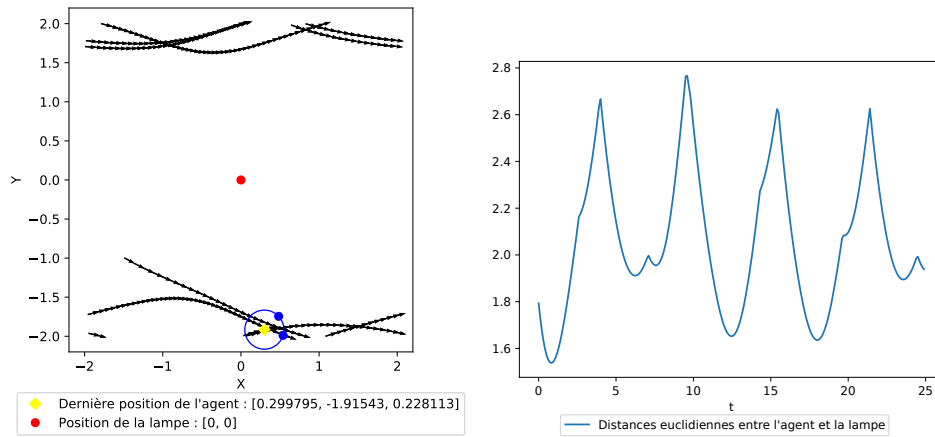
Dans le cas du comportement photophobie, les boîtes à moustache sont très étendues par rapport à celles de phototaxie, attestant de la dispersion des distances entre l’agent et la lampe (voir figure 3.9). Comme attendu, la majorité du temps l’agent est plus loin de la lampe qu’avec le comportement phototaxie : toutes les médianes et plus de la moitié des boîtes sont intégralement au dessus de 1.4 (pour rappel, pour une boîte à moustache, la boîte en elle-même représente 50% des valeurs). La figure 3.10 montre les différents comportements adoptés par l’agent. Une fois stable, il va soit osciller horizontalement (de gauche à droite ou inversement) dans le "couloir" entre  $y = 1$  et  $y = -1$  (voir l’exemple donné en figure 3.10a), soit osciller verticalement (de haut en bas ou inversement) dans le "couloir" entre  $x = 1$  et  $x = -1$  (voir l’exemple donné en figure 3.10c). Le dernier exemple (figure 3.10b) illustre un cas où il met plus de temps à se stabiliser dans un des "couloirs", ce qui résulte en une trajectoire plus variée et qui couvre davantage l’arène.

**Phototaxie sinusoïdale** Le dernier exemple reprend la phototaxie, mais en ajoutant une perturbation sinusoïdale en fonction du temps. La commande motrice forcée  $\chi$  suit les équations suivantes :

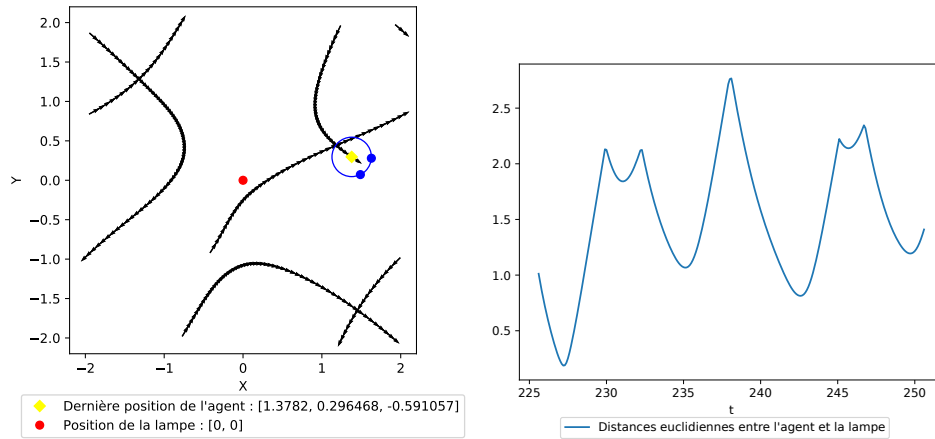
$$\chi_l = 1 - 1.5\sqrt{s_l} + \sin(2t)/2 \quad (3.25)$$

$$\chi_r = 1 - 1.5\sqrt{s_r} + \sin(2t)/2 \quad (3.26)$$

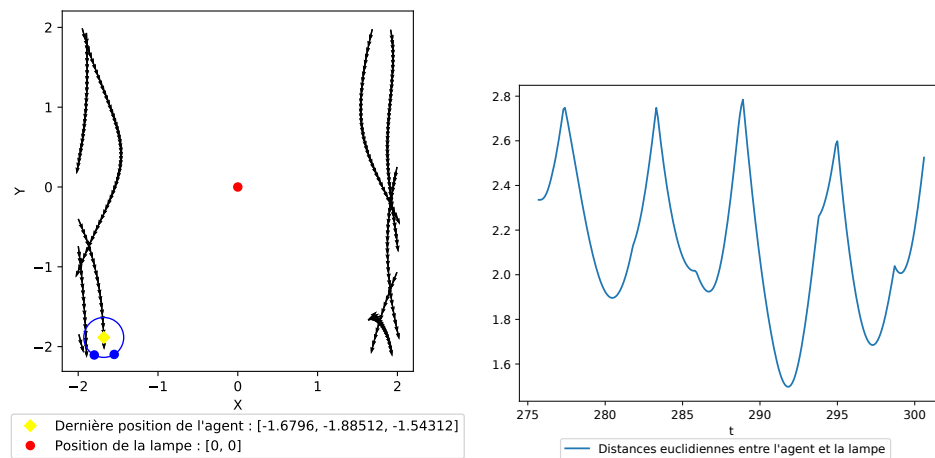




(a) Première trajectoire de l'agent pendant l'entraînement.



(b) Trajectoire 10 de l'agent.



(c) Trajectoire 12 de l'agent.

FIGURE 3.10 – Exemples de trajectoires de l'expérience sur le comportement photophobie illustrant les trois comportements de fuite observés (en fonction de la position de l'agent et de son état moteur de départ) : oscillations horizontales, oscillations verticales, ou mixte des deux le temps de se stabiliser.

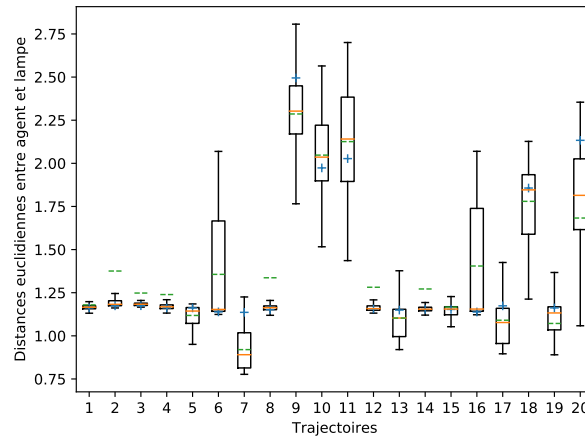
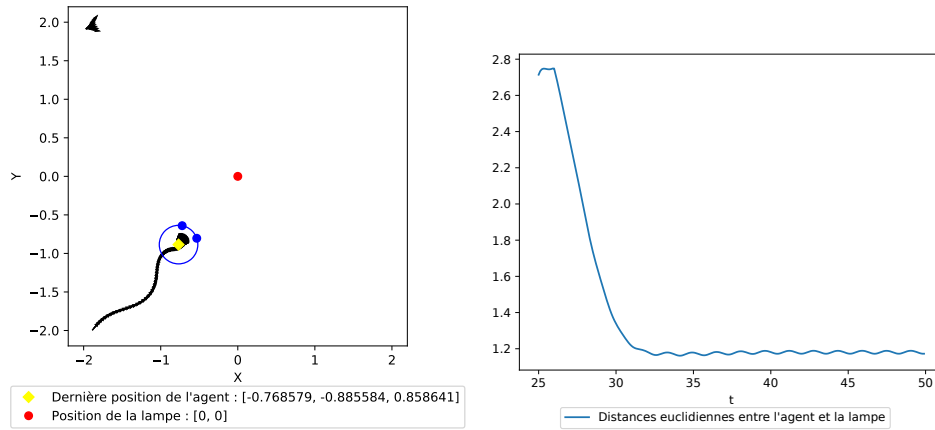


FIGURE 3.11 – Boîtes à moustache pour comparer les différences de distances entre l’agent et la lampe entre chaque trajectoire après relocalisation de l’agent pour le comportement phototaxie. Pour chaque trajectoire, la boîte est calculée à partir des 250 distances, une par pas de temps. Pour chaque boîte, le trait orange plein indique la médiane, le trait vert pointillé indique la moyenne, la croix bleue est la distance depuis la dernière position de l’agent. Les quatre premières trajectoires sont celles d’entraînement avec marche forcée. Le graphique permet de faire ressortir deux comportements : le premier qui permet à l’agent d’aller se stabiliser à une distance entre 1.15 et 1.2 de la lampe, et le second où il reste plus loin de la lampe, *a priori* sans stabiliser car les boîtes ressemblent plus à celles de photophobie (larges et avec la distance de la dernière position loin de la médiane) qu’à celle de phototaxie (petites et avec la dernière distance très proche de la médiane). Ce second comportement n’a pas été appris pendant l’entraînement et est l’agent qui fait des ronds plutôt que de se rapprocher de la lampe.

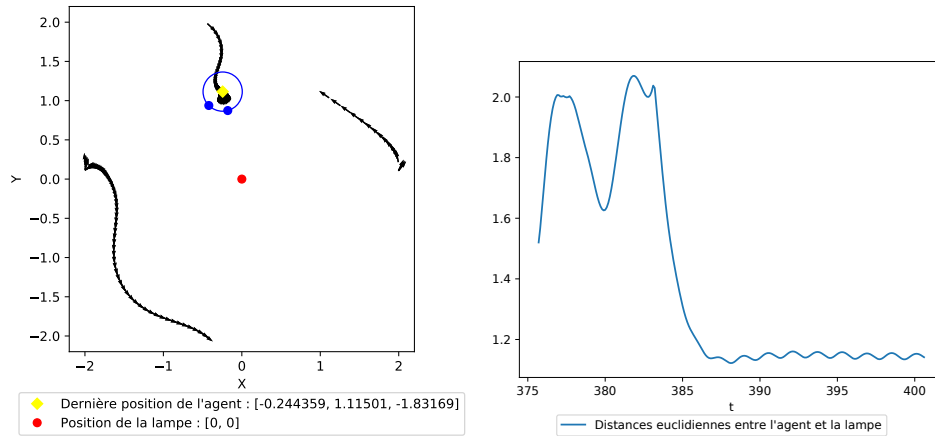
Par rapport au comportement phototaxie, l’agent va aussi chercher à se rapprocher de la lampe et s’arrêter à une certaine distance, mais en plus il va onduler et ne sera jamais à l’arrêt complet. En fin de trajectoire, il pivote toujours légèrement sur lui-même de gauche à droite puis de droite à gauche. En plus de ce comportement, un autre émerge dans les trajectoires hors entraînement : l’agent fait des ronds loin de la lampe, en bordure d’arène. Les boîtes à moustaches permettent d’identifier rapidement les trajectoires concernées (voir figure 3.11). En effet, il y a deux catégories visibles. La première catégorie contient les boîtes similaires à celles du comportement phototaxie, qui sont plus proche de la lampe à cause de la perturbation sinusoïdale (les dernières positions de l’agent sont aux alentours de 1.2). La seconde regroupe les boîtes plus distantes de la lampe (au delà de 1.5) qui résultent du comportement émergent. Des exemples de ces trajectoires sont visibles sur la figure 3.12. Le comportement émergent vient d’un équilibre local dans la dynamique de l’espace sensorimoteur qui s’est créé avec les relocalisations aléatoires, relancer cette expérience ne le produirait pas nécessairement. Inversement, il pourrait émerger d’autres comportements, comme le cercle centré sur la lampe (voir figure 3.13).

### 3.3 Exemples de combinaisons de comportements

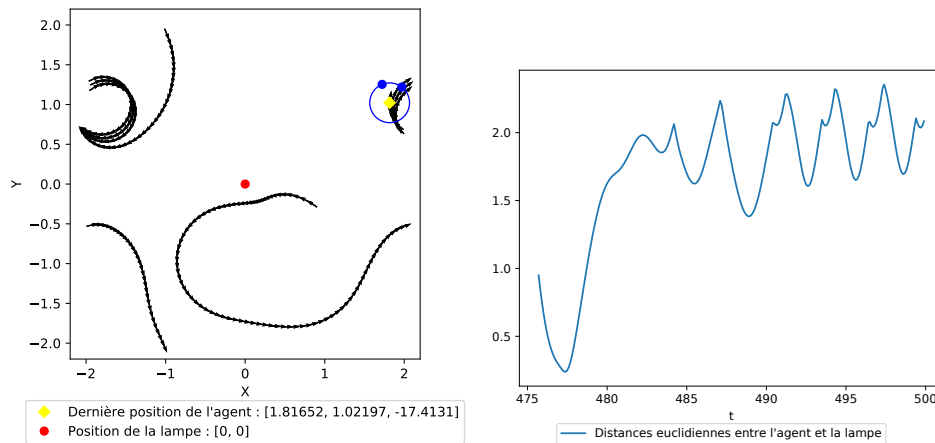
Cette partie présente nos premières expériences "originales" avec l’IDSM où l’objectif est dans un premier temps de voir comment deux comportements "préalablement acquis" interagissent entre eux et dans un second temps s’il est possible d’observer un nouveau comportement (dans



(a) Seconde trajectoire de l'agent pendant l'entraînement.



(b) Trajectoire 16 de l'agent.



(c) Trajectoire 20 de l'agent.

FIGURE 3.12 – Exemples de trajectoires de l'expérience sur le comportement phototaxie sinusoïdale. 3.12a et 3.12b montrent le comportement attendu de l'agent : serpenter vers la lampe et se stabiliser à une certaine distance d'elle (entre 1.15 et 1.2). Par contre, 3.12c montre un comportement émergent de la dynamique des nœuds : l'agent se met à faire des ronds pour le temps restant de la trajectoire.

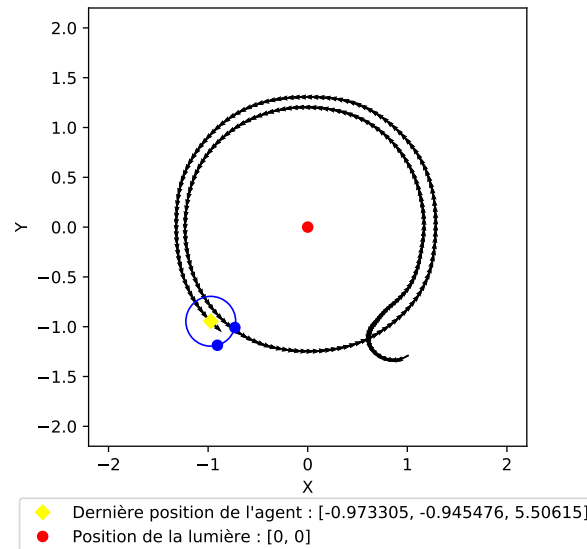


FIGURE 3.13 – Autre exemple de comportement émergent provenant d'une autre série de l'expérience avec le comportement phototaxie sinusoïdale. L'agent tourne autour de la lampe sans s'en rapprocher, au contraire il s'en éloigne petit à petit.

$\Delta t$	$k_t$	$k_d$	$k_\omega$	$\tau$	$\mu$
0.1	1	1000	0.0025	10	$\in [-\infty; +\infty]$

TABLE 3.2 – Valeurs des paramètres et options utilisées pour les exemples de combinaisons de comportements (section 3.3)

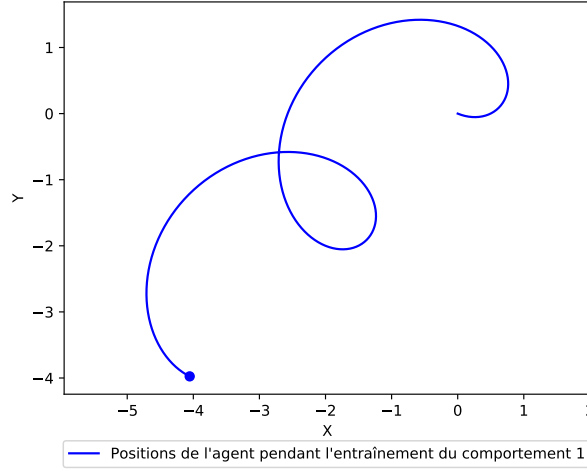


FIGURE 3.14 – Trajectoire de l’agent pendant le premier entraînement. Le point est la dernière position de l’agent.

le sens non appris au préalable) lorsque l’agent est soumis à la dynamique résultante d’une telle combinaison. Pour cela, nous reprenons un agent simple doté de deux moteurs et d’un senseur. L’agent commence par apprendre deux comportements via des entraînements en marche forcée, puis après activation de tous les nœuds nous observons le comportement résultant du contrôle de l’agent par l’IDSM. Il y a trois expériences différentes. Dans la première, le senseur n’est pas utilisé et les entraînements sont faits successivement dans le même espace sensorimoteur. Dans la seconde, le senseur n’est pas utilisé et les entraînements sont faits successivement mais dans des espaces sensorimoteurs différents qui sont fusionnés après le second entraînement : les nœuds du premier entraînement n’influent pas sur la densité et leurs poids ne sont pas modifiés lors du second entraînement. Enfin, la troisième expérience utilise le senseur en plus des espaces sensorimoteurs différents pour l’entraînement. Le premier entraînement se déroule avec un senseur forcé à  $s = 0$ , tous les nœuds du premier entraînement sont donc sur le plan de l’espace<sup>1</sup> où  $s = 0$ . Pour le second entraînement,  $s = 1$ . L’agent est ensuite relocalisé sur le plan de l’espace souhaité à l’aide d’une valeur senseur forcée  $s \in [0; 1]$ . L’expérience est relancée pour chaque plan observé. La séparation des espaces sensorimoteurs pour les entraînements n’est conservé ici que pour le gel des poids des nœuds du premier entraînement, la distance entre les plans  $s = 0$  et  $s = 1$  est suffisamment grande pour que ces nœuds n’influencent plus assez la densité pendant la création de nœuds du second entraînement.

Les commandes motrices du premier entraînement sont :

$$m_1(t) = 0.75 \cos\left(\frac{2\pi t}{10}\right) - 0.2 \quad (3.27)$$

$$m_2(t) = 0.75 \sin\left(\frac{2\pi t}{10}\right) - 0.2 \quad (3.28)$$

Le comportement résultant est que l’agent va dans une direction en bas à gauche de l’environnement en effectuant de petites boucles orientées en bas à droite (voir figure 3.14).

1. La fonction de transformation  $F$  applique l’identité pour les senseurs, nous utilisons, par abus de langage,  $s$  à la place de  $F(s)$  pour référencer la valeur de la composante sensorielle de l’état sensorimoteur.

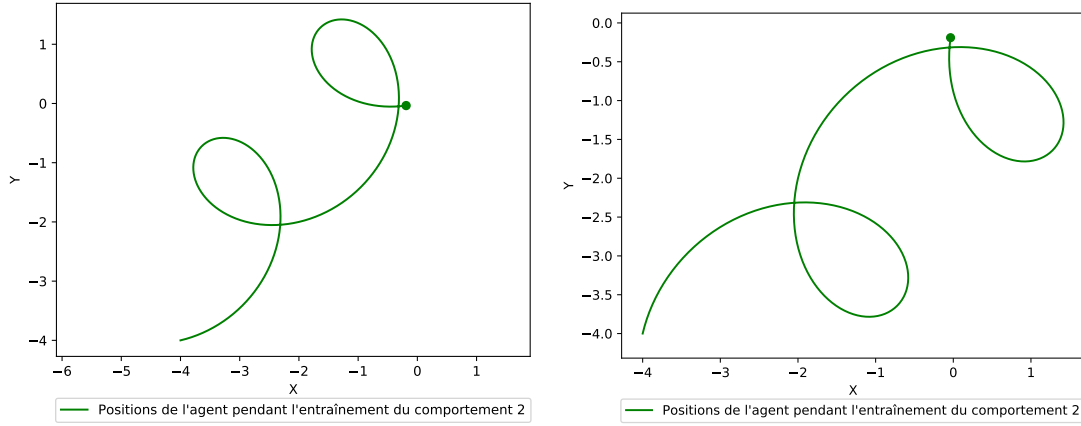


FIGURE 3.15 – Trajectoires de l’agent pendant le second entraînement. Le premier graphique correspond au motif  $P1$ , le second au motif  $P2$ . Le point est la dernière position de l’agent.

Pour le second entraînement, nous expérimentons deux motifs (*pattern*)  $P1$  et  $P2$ .

$$P1 \begin{cases} m_1(t) = 0.75 \cos\left(\frac{2\pi t}{10}\right) + 0.2 & (3.29) \\ m_2(t) = 0.75 \sin\left(\frac{2\pi t}{10}\right) + 0.2 & (3.30) \end{cases}$$

$$P2 \begin{cases} m_1(t) = 0.75 \sin\left(\frac{2\pi t}{10}\right) + 0.2 & (3.31) \\ m_2(t) = 0.75 \cos\left(\frac{2\pi t}{10}\right) + 0.2 & (3.32) \end{cases}$$

Le comportement de l’agent associé au motif  $P1$  est qu’il va dans une direction en haut à droite en effectuant de petites boucles orientées en haut à gauche. Avec  $P2$ , l’agent va aussi en haut à droite mais en conservant les mêmes boucles que pendant le premier entraînement, orientées en bas à droite. La figure 3.15 illustre les deux comportements.

### 3.3.1 Entraînements successifs

Dans cette première expérience sur les comportements combinés avec un IDSM, les deux entraînements se succèdent dans le même espace sensorimoteur, sans utiliser le senseur. Les trajectoires dans l’espace sensorimoteur s’intersectent en deux points, ce qui résulte en deux écarts dans les nœuds résultant du second entraînement par rapport à ceux du premier entraînement (voir figure 3.16 et 3.17). En effet, lors du second entraînement, la densité en ces zones est trop élevée pour créer de nouveaux nœuds. Par contre, les poids des nœuds du premier entraînement s’affaiblissent dans le temps durant le second entraînement, y compris ceux sur la trajectoire du second entraînement même si au passage de l’agent ils sont légèrement renforcés. Une fois tous les nœuds actifs, juste avant de laisser les commandes de l’agent à l’IDSM, tous les poids des nœuds du premier entraînement sont plus faibles que les poids des nœuds du second entraînement.

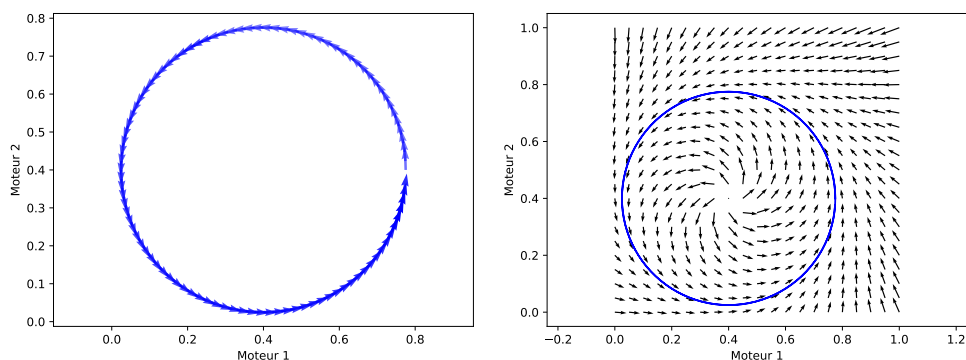
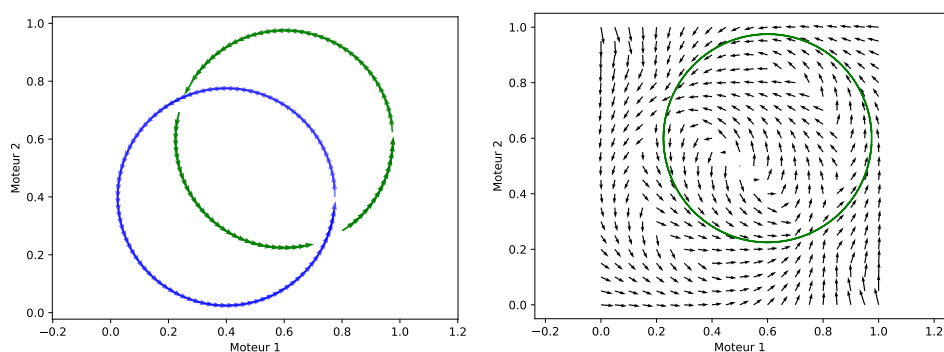
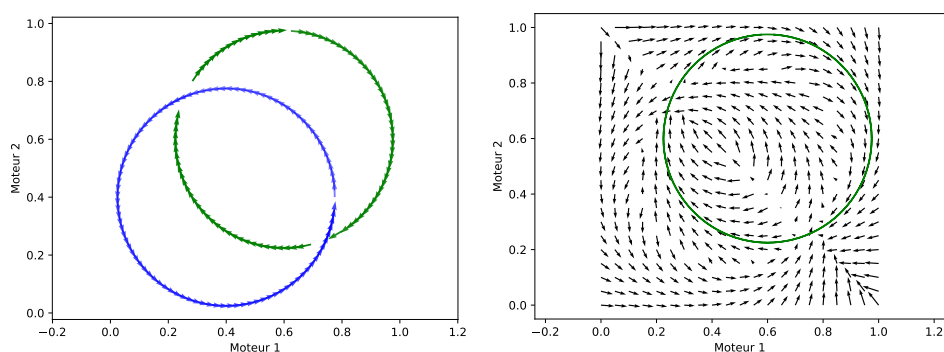


FIGURE 3.16 – Nœuds et influences à la fin du premier entraînement.



(a) Nœuds et dynamique à la fin du second entraînement avec le motif  $P1$ .



(b) Nœuds et dynamique à la fin du second entraînement avec le motif  $P2$ .

FIGURE 3.17 – Nœuds et dynamiques à la fin du second entraînement, pour les deux motifs, lorsque celui est dans le même espace sensorimoteur et à même valeur sensorielle (ici  $s = 0$ ) que le premier entraînement. Les nœuds "hérités" du premier entraînement sont en bleus, ceux du second sont en vert. Dans les graphiques à droite, la trajectoire suivie par l'agent est le cercle vert. Nous pouvons voir que pour les deux motifs, seuls les nœuds du premier entraînement (bleus) sont présents à aux intersections des trajectoires, la densité de nœuds étant trop haute pour créer de nouveaux nœuds lors du passage de l'agent pendant le second entraînement.

Pour les deux motifs, et pour nos deux points de départ de l'agent dans l'espace sensorimoteur, les résultats montrent que l'agent rejoint la trajectoire motrice du premier entraînement (voir les figures 3.18 et 3.19). Lors du premier entraînement, l'IDSM a créé des nœuds qui impactent la densité et limite fortement la création de nouveaux nœuds lors du second entraînement, en particulier aux régions de recoupement des trajectoires sensorimotrices. Les nœuds créés pendant le second entraînement ont beau avoir des poids plus élevés (car plus récents), ces nœuds sont moins nombreux et donc impactent peu la trajectoire de l'agent. Par conséquent, le comportement de l'agent dirigé par l'IDSM devient assez rapidement et durablement celui du premier entraînement : il se dirige en bas à gauche en formant de petites boucles orientées en bas à droite.

### 3.3.2 Entraînements séparés

Dans cette seconde expérience, les entraînements ont lieu dans des espaces sensorimoteurs séparés. L'idée est de supprimer leurs avantages respectifs afin que leurs dynamiques soient équivalentes. Ainsi, les nœuds du premier entraînement ne saturent plus les intersections lors du second entraînement, et les poids des nœuds du premier entraînement ne s'affaiblissent plus lors du second entraînement. Le résultat du second entraînement pour les deux motifs est visible sur la figure 3.20 : le nombre de nœuds, leur disposition et la dynamique qui en découle sont similaires à ceux du premier entraînement (voir figure 3.16). La différence dans la combinaison des dynamiques est peu visible pour le motif *P1*, entre les figures 3.18a et 3.21a : les flèches les plus proches des intersections sont plus orientées entre les trajectoires que le long de la trajectoire bleue du premier comportement. Cette différence est davantage visible pour le motif *P2*, entre les figures 3.19a et 3.22a : dans la seconde figure, la dynamique globale semble vraiment symétrique par rapport à la diagonale.

Les résultats sont très différents de ceux de la section 3.3.1. Dans tous les cas testés, l'agent va suivre un mélange des trajectoires des deux entraînements de manière stable, ce qui conduit en un comportement différent de ceux des entraînements. Pour le motif *P1* (voir figure 3.21), selon la position de relocalisation, l'agent va suivre le périmètre de l'intersection des deux cercles ou celui de leur union. Dans le premier cas, le comportement résultant est une petite ellipse qui est la combinaison des boucles. Dans le second, l'agent effectue une grande ellipse qui est la combinaison des arcs.

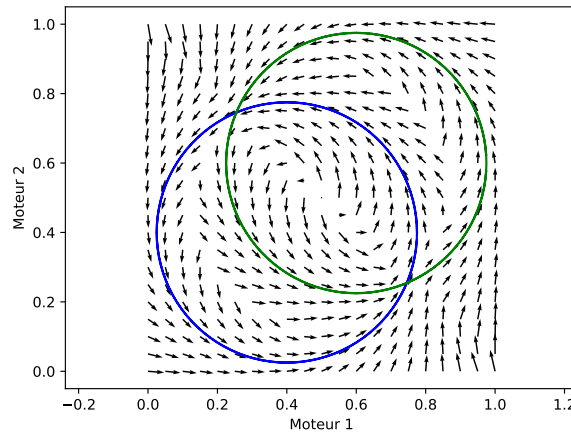
Pour le motif *P2* (voir figure 3.22), pour les deux relocalisations, l'agent suit le croissant constitué des parties en haut à droite des cercles. Cela est dû aux poids des nœuds aux intersections, plus ou moins renforcés selon les trajectoires d'entraînements (voir les nœuds de la figure 3.16 et ceux de la figure 3.20b). Pour l'intersection du haut, les nœuds les plus récents et ayant un poids plus forts sont les nœuds du second entraînement. Pour celle du bas, c'est l'inverse, il s'agit de ceux du premier entraînement. Le comportement résultant de cette trajectoire en croissant est un enchaînement d'arcs sans boucles et il conserve la direction du second entraînement en allant vers le haut à droite.

La séparation des espaces sensorimoteurs pour les entraînements est donc une façon efficace de faire émerger un comportement dans ce cadre simple.

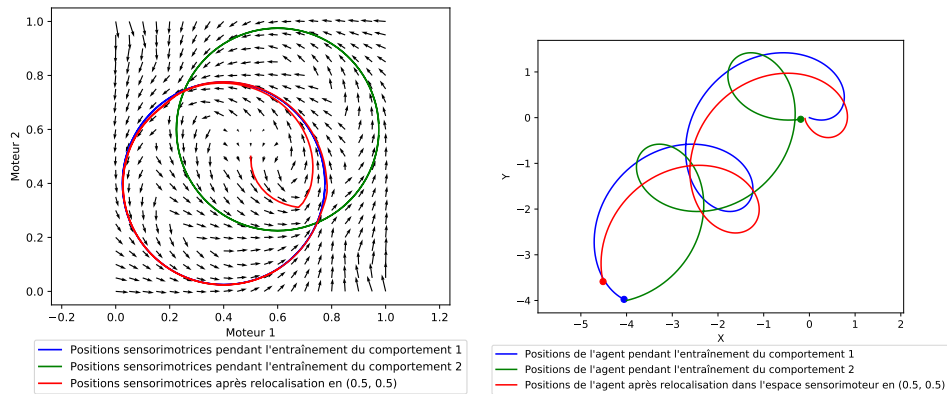
### 3.3.3 Comportements dépendants d'un senseur

L'objectif de cette dernière expérience est d'observer le comportement résultant de l'agent lorsque les deux entraînements ont lieu à des valeurs différentes du senseur ( $s = 0$  pour le premier

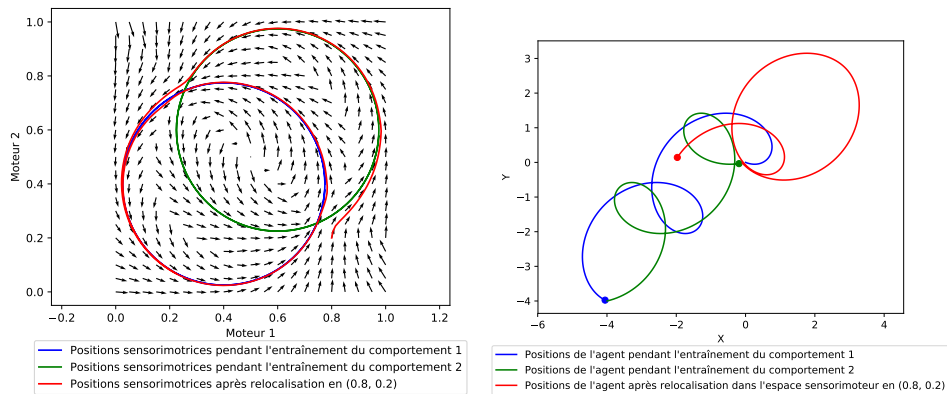




(a) Dynamique dans l'espace sensorimoteur avec tous les nœuds actifs.

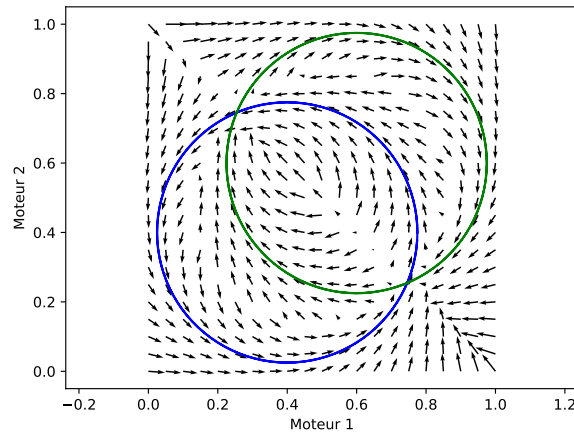


(b) Dynamique et trajectoire obtenues après relocalisation en (0.5, 0.5).

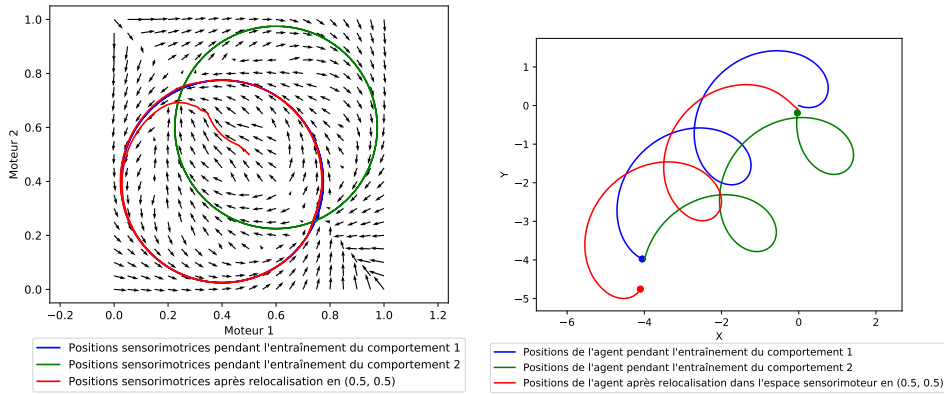


(c) Dynamique et trajectoire obtenues après relocalisation en (0.8, 0.2).

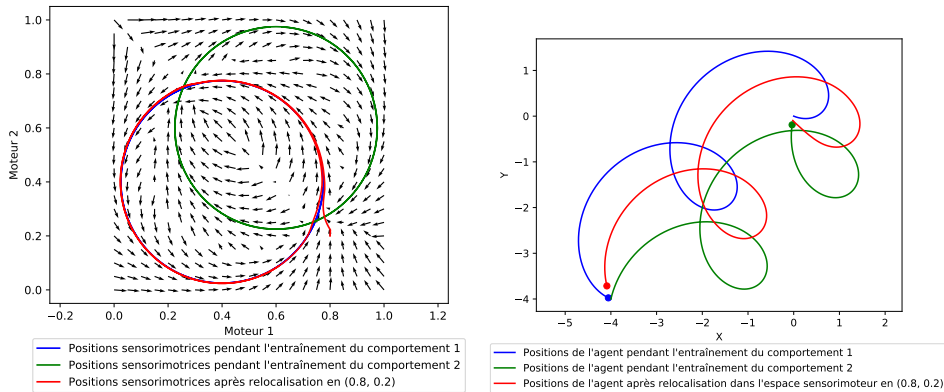
FIGURE 3.18 – Résultat pour le motif  $P1$  lorsque les deux entraînements ont lieu dans le même espace sensorimoteur. Dans l'espace sensorimoteur, l'agent commence par aller vers la trajectoire du second entraînement mais finit toujours par suivre la dynamique du premier entraînement, plus rapidement lorsqu'il est relocalisé en (0.5, 0.5). Il suit cette dynamique jusqu'à la fin de l'expérience avec très peu d'écart par rapport à la marche forcée. Son comportement stable est celui du premier entraînement (en bleu) : l'agent se dirige vers en bas à gauche en formant des boucles orientées en bas à droite.



(a) Dynamique dans l'espace sensorimoteur avec tous les nœuds actifs.



(b) Dynamique et trajectoire obtenues après relocalisation en (0.5, 0.5).



(c) Dynamique et trajectoire obtenues après relocalisation en (0.8, 0.2).

FIGURE 3.19 – Résultat pour le motif  $P2$  lorsque les deux entraînements ont lieu dans le même espace sensorimoteur. Dans l'espace sensorimoteur, contrairement au motif  $P1$ , le départ de l'agent est moins tranché, il n'est pas dirigé directement sur l'une ou l'autre des trajectoire d'entraînement mais oscille légèrement entre les deux. Une fois assez proche par contre il rejoint directement la dynamique résultant du premier entraînement et la suit jusqu'à la fin de l'expérience. Le comportement résultant est celui du premier entraînement : l'agent se dirige en bas à gauche (avec le motif  $P2$  les boucles sont similaires pour les deux entraînements).

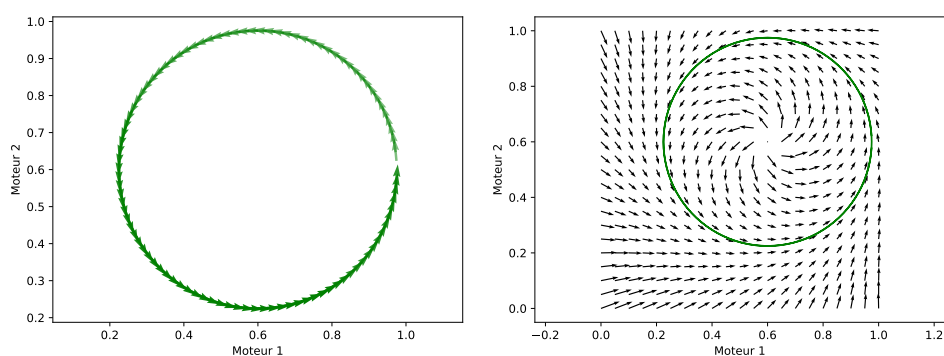
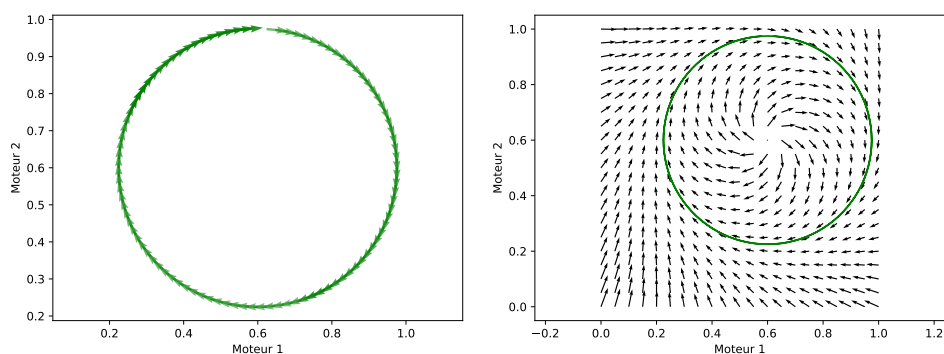
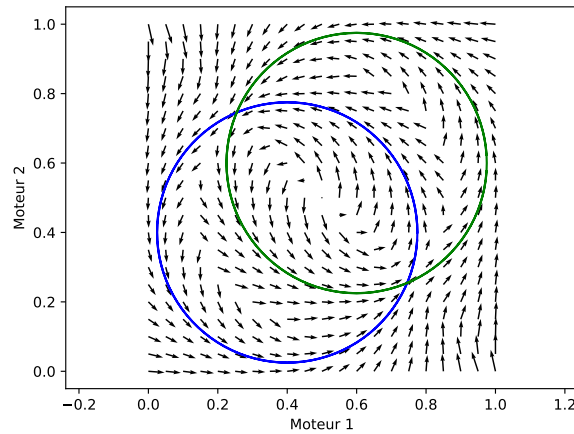
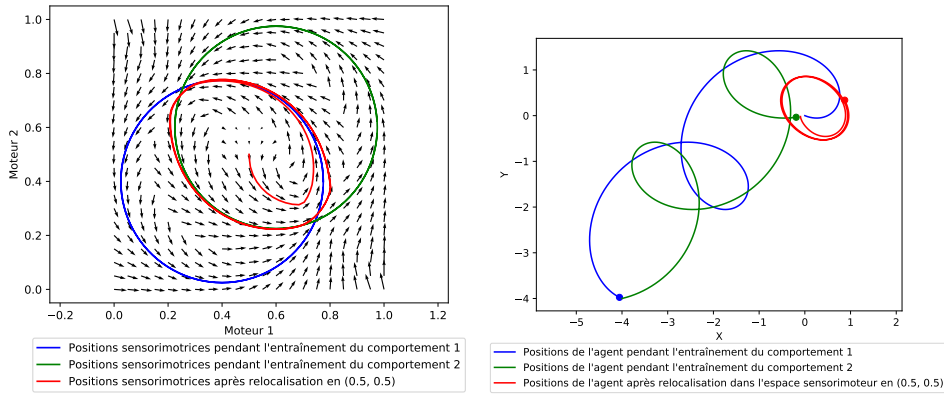
(a) Nœuds et dynamique à la fin du second entraînement avec le motif  $P1$ .(b) Nœuds et dynamique à la fin du second entraînement avec le motif  $P2$ .

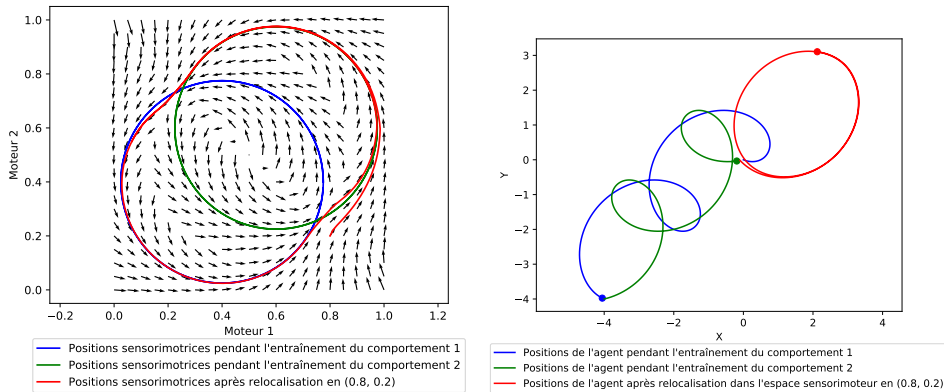
FIGURE 3.20 – Nœuds et dynamique à la fin du second entraînement, pour les deux motifs, lorsque cet entraînement est dans un espace sensorimoteur différent de celui du premier entraînement mais toujours avec la même valeur sensorielle (ici  $s = 0$ ). Nous pouvons voir que les nœuds couvrent bien les intersections contrairement à ceux de la figure 3.17, et que la dynamique est symétrique à celle du premier entraînement (voir figure 3.16).



(a) Dynamique dans l'espace sensorimoteur avec tous les nœuds actifs.

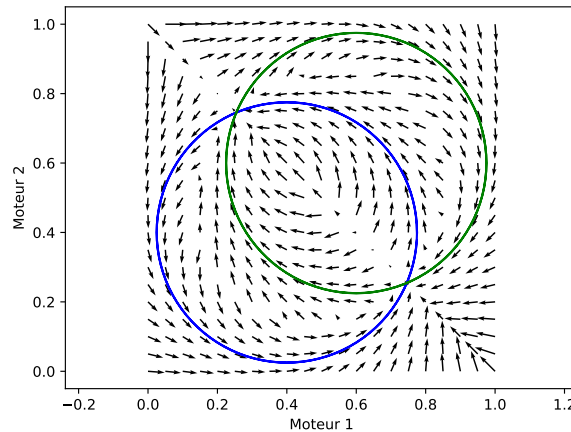


(b) Dynamique et trajectoire obtenues après relocalisation en  $(0.5, 0.5)$ .

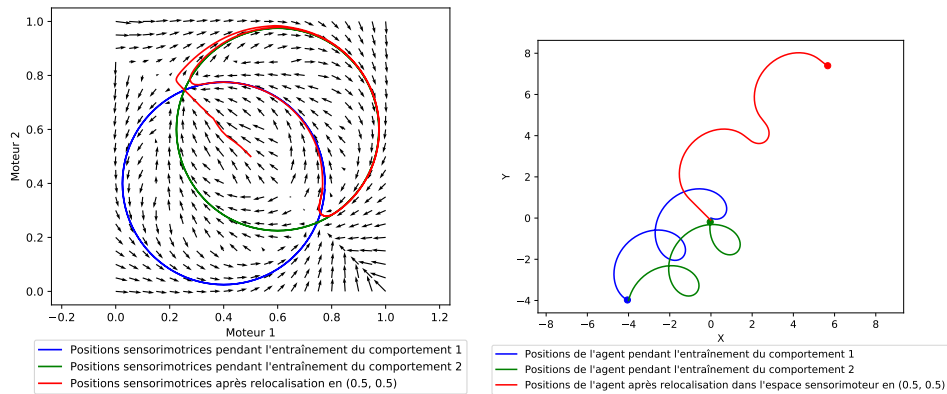


(c) Dynamique et trajectoire obtenues après relocalisation en  $(0.8, 0.2)$ .

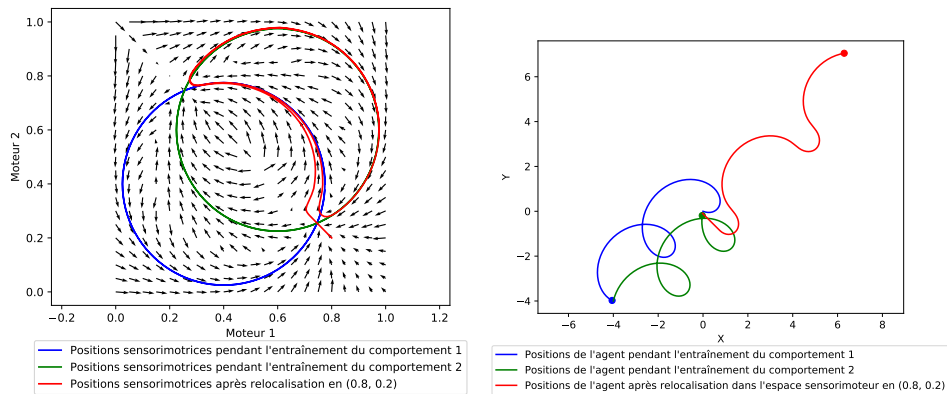
FIGURE 3.21 – Résultat pour le motif  $P1$  lorsque les entraînements utilisent des espaces sensorimoteurs distincts. Dans l'espace sensorimoteur, l'agent alterne entre les trajectoires de chaque entraînement, soit en restant dans la partie interne lorsque la relocalisation est en  $(0.5, 0.5)$ , soit au contraire en restant sur la partie externe avec la relocalisation en  $0.8, 0.2$ ). Cela se traduit par un comportement différent de ceux des entraînements. Pour la première relocalisation l'agent effectue une petite ellipse qui est la combinaison des boucles. Pour la seconde relocalisation l'agent effectue une grande ellipse qui est la combinaison des arcs.



(a) Dynamique dans l'espace sensorimoteur avec tous les nœuds actifs.



(b) Dynamique et trajectoire obtenues après relocalisation en  $(0.5, 0.5)$ .



(c) Dynamique et trajectoire obtenues après relocalisation en  $(0.8, 0.2)$ .

FIGURE 3.22 – Résultat pour le motif  $P2$  avec espaces sensorimoteurs distincts pendant les entraînements. Les deux relocalisations conduisent au même état stable en forme de croissant. Pour l'intersection du haut, les nœuds les plus récents et ayant donc un poids plus forts sont les nœuds du second entraînement ; pour celle du bas, c'est l'inverse, il s'agit de ceux du premier entraînement (voir les nœuds de la figure 3.16 et ceux de la figure 3.20b). Le comportement de l'agent est dans les deux cas un enchaînement d'arcs sans boucles et il conserve la direction du second entraînement.

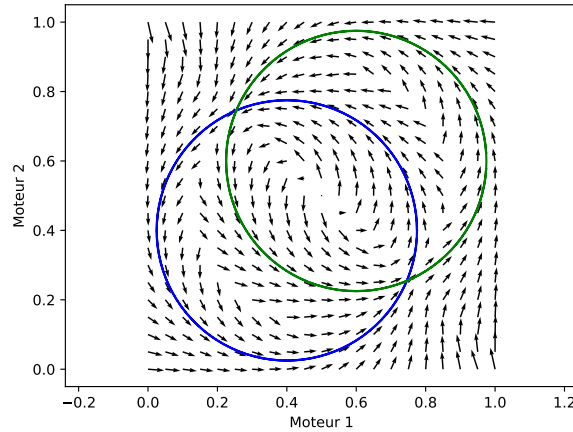
entraînement,  $s = 1$  pour le second). En plus de permettre une combinaison de comportements sans les conflits de densité vu dans la section 3.3.1, le senseur peut alors servir de levier pour changer d'un comportement à l'autre. Cette étude est préliminaire aux études du chapitre suivant sur l'habituation et la sensibilisation spécifique au site qui vont être produites par l'activation et la désactivation d'un stimulus.

Les résultats sont observés pour la valeur sensorielle  $s = 0.5$ , pour comparer la combinaison de comportements avec celle de la section 3.3.2. Pour les deux motifs (voir figures 3.23 et 3.24), les trajectoires sont respectivement similaires à celles des figures 3.21 et 3.22, avec comme différence notable qu'il y a dans chaque cas une perturbation dans les trajectoires motrices : petite boucle vers le centre dans le cas de la figure 3.23b, cassure dans la trajectoire en se rapprochant du centre vers  $[0.35, 0.05]$  pour la figure 3.23c, vers  $[0.75, 0.45]$  pour la figure 3.24b et vers  $[0.95, 0.5]$  pour la figure 3.24c. Dans tous les cas, il s'agit exactement du moment où le premier nœud créé après relocalisation s'active. Ce nœud a une forte influence à ce moment car il est plus proche de l'agent que les nœuds d'entraînements, d'où la perturbation immédiate sur la dynamique. Avec l'activation successive des autres nœuds, l'agent retrouve un état stable prenant en compte la perturbation (3.23c, 3.24b et 3.24c) ou non (3.23b). Le comportement résultant est impacté par la brusque activation des nouveaux nœuds d'une part, mais aussi de manière durable lorsqu'il y a intégration de la perturbation dans l'état stable. Par exemple, sur du plus long terme, l'agent de la figure 3.23c continuerait de faire dévier ses ellipses vers le haut plutôt que de les faire au même endroit. Nous pouvons noter que la dynamique finale est influencée presque exclusivement par les nouveaux nœuds. Cela se voit en comparant les dynamiques finales avec la dynamique en fin d'entraînements (voir figure 3.23a et 3.24a) : les flèches noires "ignorent" les trajectoires bleue et verte, seule la rouge est prise en compte.

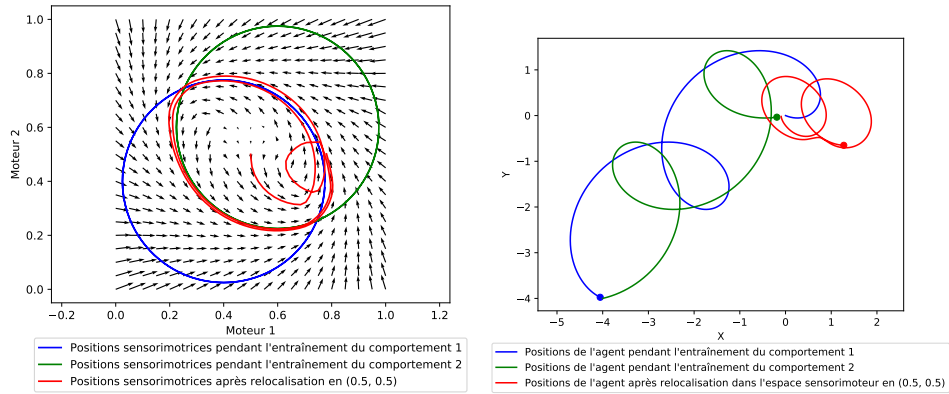
La figure 3.25 illustre que modifier la valeur du senseur permet de retrouver les comportements d'entraînement avec une perturbation similaire à celles observés en  $s = 0.5$  après l'activation des nouveaux nœuds. Cette perturbation est d'autant plus faible que la valeur du senseur se rapproche de 0 ou de 1. Un léger écart (par exemple  $s = 0.49$ ) est suffisant pour redonner l'avantage à l'une ou l'autre dynamique d'entraînement.

### 3.3.4 Bilan de ces expériences

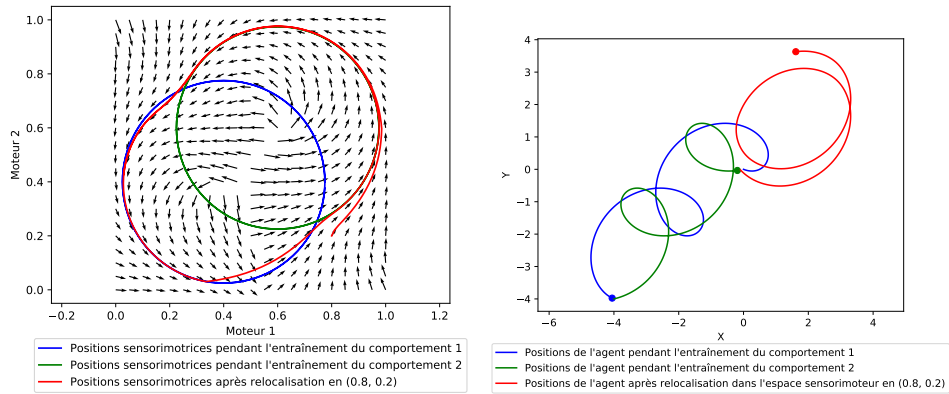
Ces trois expériences permettent d'avoir des premiers résultats sur l'interaction entre différents comportements dans un IDSM. L'agent acquiert ces comportements au préalable par marche forcée lors d'entraînements. Dans la première expérience, les nœuds du premier entraînement saturent la densité des zones traversées communément par les deux trajectoires sensorimotrices d'entraînements, ce qui empêche le second entraînement d'être complètement mémorisé. Par conséquent, les comportements observés après entraînements sont tous similaires à celui du premier entraînement seul. Au contraire, la seconde expérience montre qu'en entraînant l'agent avec des espaces sensorimoteurs différents, les entraînements permettent la création de nœuds équivalents. Une fois ces ensembles de nœuds fusionnés, l'équilibre entre eux permet alors une trajectoire motrice dépendante des deux entraînements et un comportement résultant qui est une combinaison des comportements d'entraînements. Enfin, la dernière expérience montre que les nœuds des entraînements peuvent être séparés par une dimension sensorielle, cette dimension pouvant alors servir de levier pour passer progressivement du premier comportement appris au second, avec un comportement intermédiaire qui est une combinaison des deux comportements d'entraînements.



(a) Dynamique dans l'espace sensorimoteur avec tous les nœuds actifs.

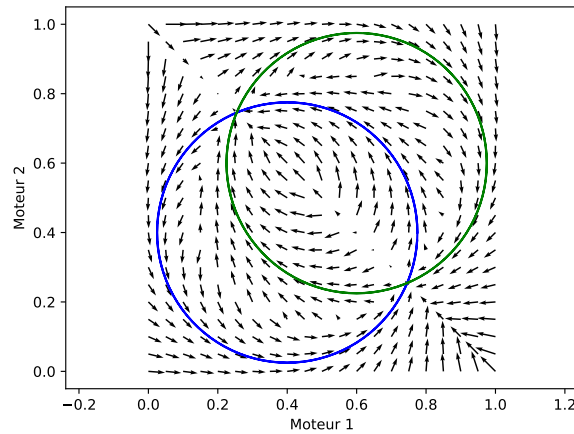


(b) Dynamique et trajectoire obtenues après relocalisation en  $(0.5, 0.5)$ .

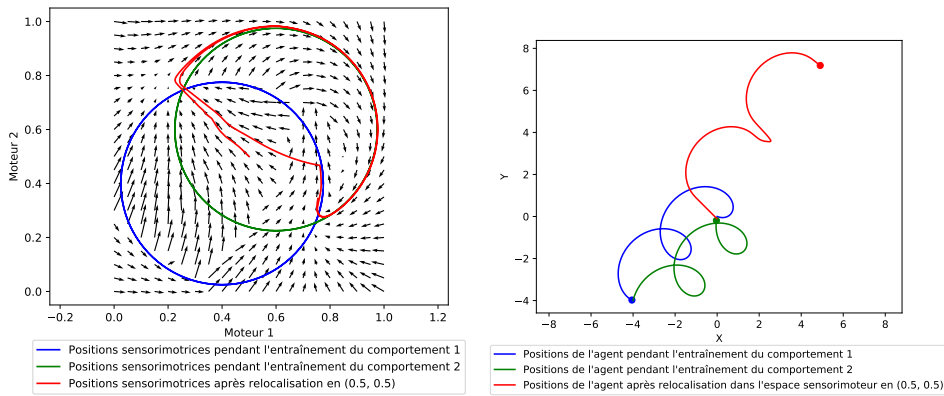


(c) Dynamique et trajectoire obtenues après relocalisation en  $(0.8, 0.2)$ .

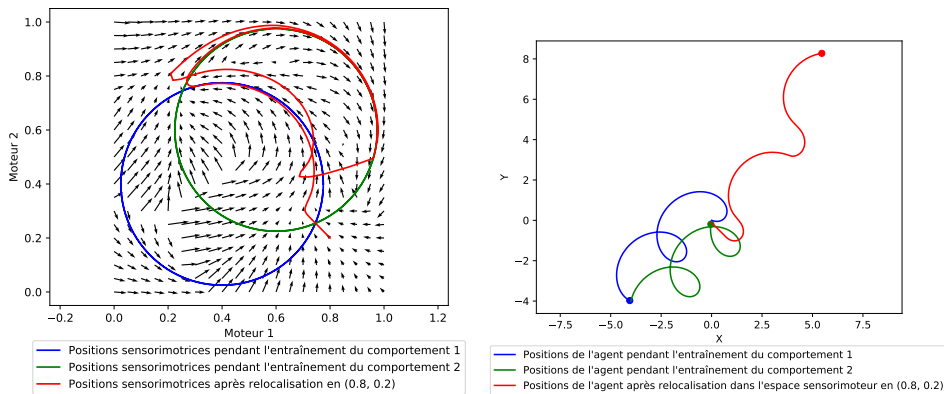
FIGURE 3.23 – Résultat pour le motif  $P1$  lorsque les entraînements utilisent des espaces sensorimoteurs distincts et sont en plus dépendant de la valeur du capteur ( $s = 0$  pour le premier entraînement en bleu,  $s = 1$  pour le second en vert). Les trajectoires sont similaires à celles de la figure 3.21, avec comme différence notable qu'il y a dans chaque cas une perturbation (petite boucle pour la relocalisation en  $(0.5, 0.5)$ , écart avec la trajectoire bleue vers  $m_2 = 0$  pour l'autre relocalisation). Dans les deux cas, cela est dû à l'activation du premier nouveau nœud qui est plus proche de l'agent que les nœuds d'entraînements. Avec l'activation successive des autres nœuds, l'agent retrouve un état stable proche de celui de la figure 3.21



(a) Dynamique dans l'espace sensorimoteur avec tous les nœuds actifs.



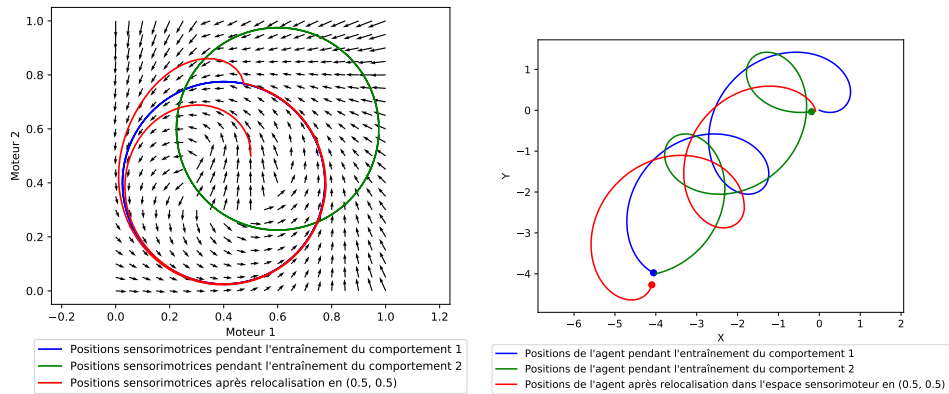
(b) Dynamique et trajectoire obtenues après relocalisation en  $(0.5, 0.5)$ .



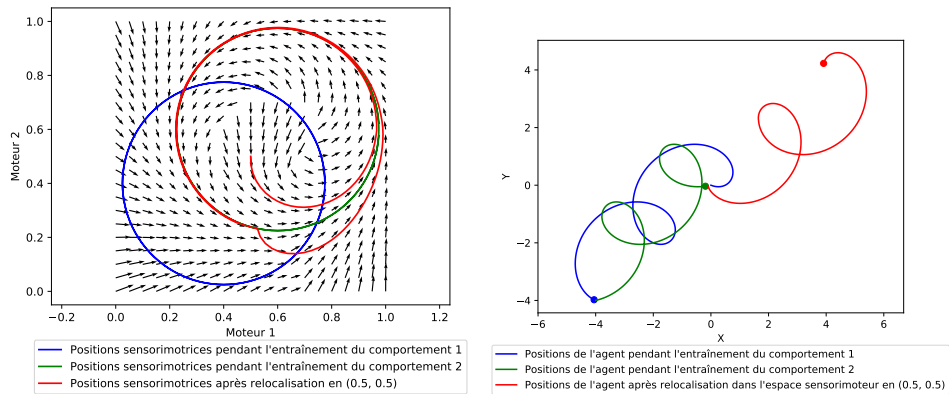
(c) Dynamique et trajectoire obtenues après relocalisation en  $(0.8, 0.2)$ .

FIGURE 3.24 – Résultat pour le motif  $P2$  lorsque les entraînements utilisent des espaces sensorimoteurs distincts et sont en plus dépendant de la valeur du senseur ( $s = 0$  pour le premier entraînement en bleu,  $s = 1$  pour le second en vert). Comme pour le motif  $P1$ , l'activation des nouveaux nœuds entraîne une perturbation brusque dans la trajectoire motrices et les états stables sont en partie modifiés par rapport à ceux de la figure 3.22. Il en va de même pour le comportement de l'agent qui dépend directement de ces trajectoires motrices.





(a) Dynamique et trajectoire après relocalisation en  $(0.5, 0.5)$  et avec le capteur forcé à  $s = 0.25$ .



(b) Dynamique et trajectoire après relocalisation en  $(0.5, 0.5)$  et avec le capteur forcé à  $s = 0.75$ .

FIGURE 3.25 – Résultats qui illustre que forcer le capteur plus d'un côté que de l'autre permet de favoriser tel ou tel entraînement et par conséquent son comportement associé. Les perturbations dans les trajectoires motrices sont dues à l'activation des nouveaux nœuds. Ces derniers sont plus proches de l'agent et avec un poids plus fort que ceux des entraînements.

Ces expériences sont une première tentative d’avoir une combinaison de comportements, elles montrent d’une part qu’un nouveau comportement peut être exprimé à partir d’interactions de dynamiques apprises via différents comportements et d’autre part que la variation d’un stimulus perçu peut permettre d’explorer ces différences dans la dynamique et les comportements associés. De plus, ces expériences ont été l’occasion d’approfondir notre compréhension du modèle IDSM et de nous familiariser avec ses différents constituants. Avant de conclure ce chapitre, nous discutons dans la prochaine section de l’effet des différents paramètres et autres composants de ce modèle.

### 3.4 Effets des paramètres de l’IDSM et autres points d’attention

Cette dernière partie du chapitre 3 a pour objectif d’illustrer plus concrètement les effets des variations dans les paramètres de l’IDSM, ainsi que de mettre en avant des points d’attention significatifs qui résultent de choix de notre part et/ou de caractéristiques propres à ce modèle.

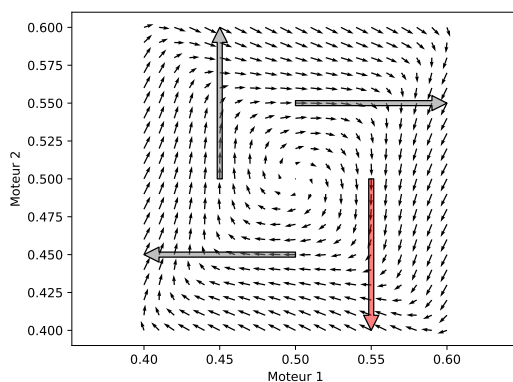
#### 3.4.1 Un seul nœud par position sensorimotrice

Lors du premier passage de l’agent à une position sensorimotrice  $\mathbf{x}_{SM}$  faible en densité de nœuds ( $\phi_N(\mathbf{x}_{SM}) < k_t$ ), un nœud  $N$  est créé avec un certain vecteur vitesse  $\mathbf{N}_v$ . Ce nœud est inamovible et non supprimable. Son vecteur vitesse  $\mathbf{N}_v$  contient l’information de la variation de la commande motrice au moment de la création du nœud  $N$ , information cohérente à ce moment précis et non modifiable par la suite. Par conséquent, il est difficile de changer la dynamique en  $\mathbf{x}_{SM}$  une fois le nœud créé. Tout futur passage de l’agent en  $\mathbf{x}_{SM}$  renforcera  $N$ , et un nouveau nœud avec un vecteur  $\mathbf{v}$  en phase avec la dynamique actuelle ne sera créé que si le poids de  $N$  est suffisamment faible. La dynamique par défaut de l’IDSM est pensée pour renforcer l’existant, pour une certaine robustesse, et par conséquent il peut être compliqué de lui apporter la plasticité nécessaire pour l’adaptation de l’agent aux changements dans son environnement. Ce premier point est illustré dans la section 3.3.1 : à l’intersection des trajectoires sensorimotrices des deux entraînements, seuls les nœuds du premier entraînement sont présents et cela conduit à une dominance de sa dynamique sur le second entraînement, et ce malgré le fait que les nœuds du second entraînement sont plus récents et ont donc un poids plus important.

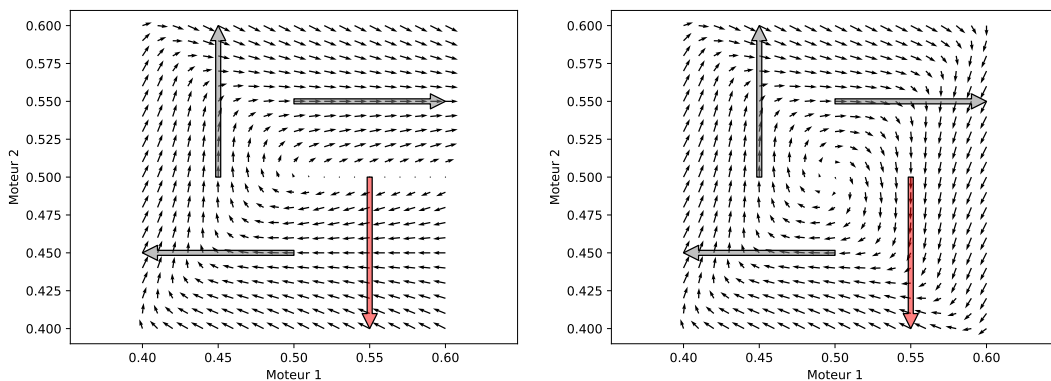
#### 3.4.2 Le poids des nœuds $N_w$

Le poids  $N_w$  du nœud  $N$  est sa composante permettant d’augmenter ou de diminuer son influence dans le temps sur la dynamique du système. Au minimum, et ce peu importe la valeur des paramètres, l’influence du nœud devient nulle puisque  $\omega(N_w) \xrightarrow{N_w \rightarrow -\infty} 0$  (voir équation 3.3). Au maximum, elle est doublée par rapport à celle que le nœud a avec un poids  $N_w = 0$  (par défaut à sa création) car  $\omega(N_w) \xrightarrow{N_w \rightarrow +\infty} 2$ . Par conséquent, un nœud gagne potentiellement davantage d’influence sur la dynamique du système via la baisse des poids des autres nœuds que par le renforcement de son propre poids. La figure 3.26 illustre cette caractéristique. Le nœud  $N$  représenté en rouge voit son poids modifié selon les différents graphiques, les trois autres nœuds restent à un poids nul. Avec un poids  $N_w = 5000$ , le facteur  $\omega(N_w)$  vaut environ 2, par conséquent l’influence de  $N$  est presque doublée par rapport aux trois autres nœuds avec leurs poids à 0, mais l’impact sur la dynamique du système n’est quasiment pas visible. À l’inverse, avec un poids  $N_w = -5000$ , le facteur  $\omega(N_w)$  vaut environ 0, son influence est quasiment inexistante. Les nœuds les plus proches (flèches horizontales sur la figure 3.26) étendent leur dynamique sur la moitié droite de l’espace où se trouve le nœud  $N$  affaibli. C’est pourquoi, plutôt que d’avoir

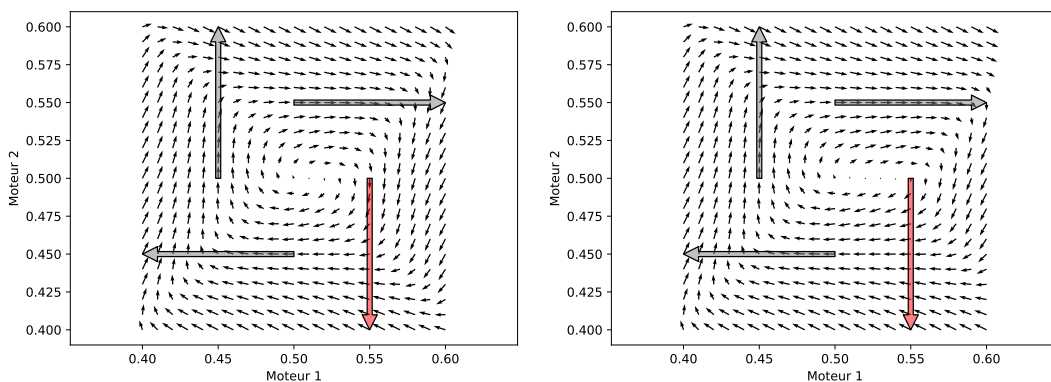
un nœud qui se renforce ou s'affaiblit après sa création, il est plus correct de visualiser un nœud qui se *maintient* ou s'affaiblit après sa création.



(a) Effet de  $N_w = 0$ .



(b) Effet de  $N_w = -5000$  à gauche et  $5000$  à droite.



(c) Effet de  $N_w = -1000$  à gauche et  $-2000$  à droite.

FIGURE 3.26 – Évolution de la dynamique avec un IDSM de deux dimensions motrices et seulement 4 nœuds. Les nœuds représentés par les flèches grises ont toujours une poids  $N_w = 0$ . Sur la figure 3.26a, la dynamique est équilibrée. Dans les cas extrêmes (figure 3.26b), le nœud rouge devient comme inexistant à gauche, contre une renforcement d'influence à peine perceptible à droite. Enfin, la figure 3.26c montre l'affaiblissement progressif du nœud rouge, ce qui se traduit par l'agrandissement de la zone d'équilibre depuis le centre vers la droite.

Un autre point d'attention sur les poids est par rapport à l'équation de leur mise à jour (équation 3.5). Cette équation n'est paramétrable qu'à travers le facteur  $d(\mathbf{N}_p, \mathbf{x}_{SM})$ , donc par le biais du paramètre  $k_d$ . En particulier, en fonction de la durée d'une expérience, la perte maximale d'une unité de poids par unité de temps peut sembler disproportionnée par rapport aux ordres de grandeurs des valeurs de poids nécessaires pour un impact significatif (voir figures 3.27 et 3.26). Par exemple, dans un cas où l'agent passe 2000 unités de temps loin du nœud, ce dernier aurait un poids d'environ  $-2000$  et serait toujours influent, comme nous pouvons le voir sur la figure 3.26c, même proche d'autres nœuds ayant un poids supérieur (leurs poids valant 0).

### 3.4.3 Hyperparamètre $k_\omega$ pour la fonction de poids $\omega(N_w)$

$\omega(N_w)$  est la fonction qui détermine l'impact du poids d'un nœud sur les fonctions de densité  $\phi_A$  et  $\phi_N$ , ainsi que sur les vecteurs de vitesse  $\mathbf{V}$  et d'attraction  $\mathbf{A}$  (voir équation 3.3 et le second graphique de la figure 3.3). Le paramètre  $k_\omega$  permet d'atteindre des valeurs proches de ces limites avec un intervalle de poids plus ou moins étendu. Par exemple, sur le second graphique de la figure 3.3, la courbe est tracée pour l'intervalle  $[-2000, 2000]$  avec  $k_\omega = 0.0025$ . En suivant l'équation 3.3, si nous voulons les mêmes influences maximale et minimale pour des poids deux fois inférieurs, il faut multiplier  $k_\omega$  par deux. Les figures 3.27, 3.28 et 3.29 donnent trois visualisations de variations de  $k_\omega$ . Avec un  $k_\omega$  plus élevé, la variation de poids est plus significative, un nœud va pouvoir plus rapidement doubler son influence ou être négligeable dans la dynamique du système. À l'inverse, un faible  $k_\omega$  va rendre le système moins sensible aux variations de poids, il faudra que celles-ci soient plus importantes pour avoir un impact similaire.

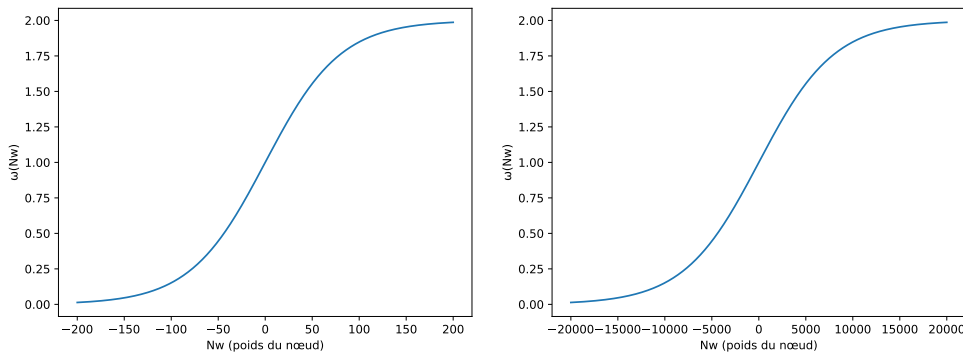


FIGURE 3.27 –  $\omega(N_w)$  en fonction de  $N_w$  avec  $k_\omega = 0.025$  à gauche et  $k_\omega = 0.00025$  à droite. Les courbes sont identiques (y compris avec celle du second graphique de la figure 3.3), c'est l'ordre de grandeur de l'axe des abscisses qui change.

### 3.4.4 Hyperparamètre $k_d$ pour la fonction de distance $d(\mathbf{N}_p, \mathbf{x}_{SM})$

Le paramètre  $k_d$  intervient dans le calcul de la fonction  $d(\mathbf{N}_p, \mathbf{x}_{SM})$  entre l'état sensorimoteur de l'agent et un nœud  $N$  (voir l'équation 3.4 et le graphique de gauche de la figure 3.3). Augmenter  $k_d$  réduit la distance d'influence d'un nœud, et inversement (voir la figure 3.30). La fonction  $d$  étant aussi utilisée dans le calcul de mise à jour des poids (voir équation 3.5), un  $k_d$  plus faible permet aussi d'augmenter le poids d'un nœud à une distance euclidienne plus grande entre lui et l'agent. La figure 3.31 illustre trois valeurs de  $k_d$  (30, 2000 et 10000) dans le cas simple avec le comportement cercle de la section 3.2.1. Dans le cas de  $k_d = 30$ , le renforcement de la densité provoque une saturation ( $\phi_N \geq k_t$ ) précoce en fin de trajectoire, ce qui implique l'impossibilité de créer de nouveaux nœuds sur cette partie de trajectoire. Ce manque d'information dans la

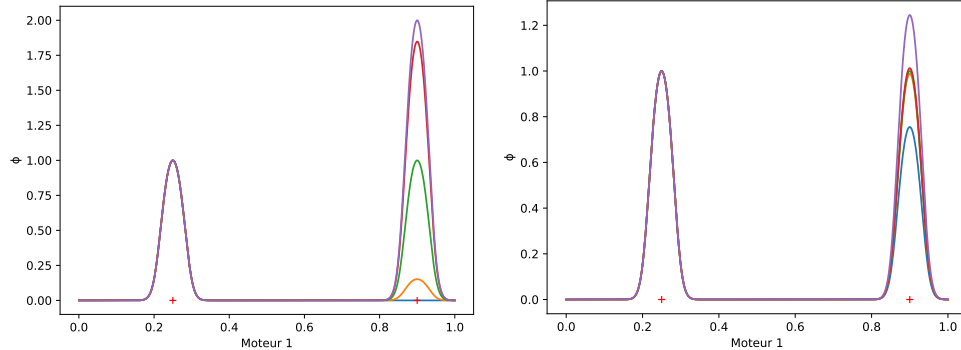


FIGURE 3.28 – Affichage de la densité  $\phi_N$  pour un IDSM en une dimension, avec  $k_\omega = 0.025$  à gauche et  $k_\omega = 0.00025$  à droite. Le poids du nœud en 0.9 vaut  $-2000$  (en bleu),  $-100$  (orange),  $0$  (vert),  $100$  (rouge) et  $2000$  (violet). Le poids du premier nœud en 0.25 reste à 0. Avec un faible  $k_\omega$ , un nœud sera moins renforcé et moins affaibli pour une même variation de poids qu’avec un  $k_\omega$  plus élevé.

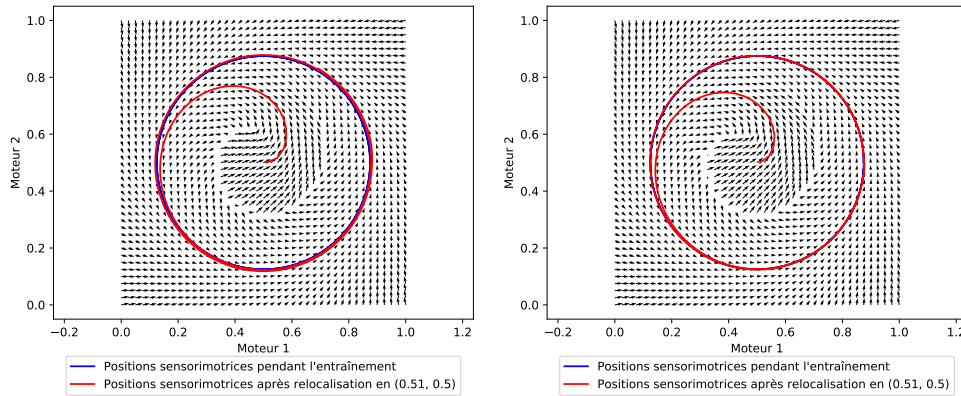


FIGURE 3.29 – Dynamiques en fin d’expérience avec  $k_\omega = 0.25$  à gauche et  $k_\omega = 0.00025$  à droite. Le choix de  $k_\omega = 0.25$  est nécessaire pour voir un impact, les poids ne pouvant pas descendre en dessous de  $-40$  ( $-1$  par unité de temps dans le pire des cas, voir l’équation 3.5). Le graphique de droite est identique à celui de la figure 3.5c. Sur le graphique de gauche, la trajectoire rouge s’éloigne légèrement de la trajectoire d’entraînement (en bleu). La densité devient assez faible entre chaque passage de l’agent pour qu’il recrée un nouveau nœud, renforçant le décalage.

dynamique empêche l'IDSM de guider parfaitement l'agent sur la trajectoire motrice en cercle. Le comportement résultant est un cercle glissant. Le cas intermédiaire avec  $k_d = 2000$  est similaire au cas avec  $k_d = 1000$  de la section 3.2.1 (voir figure 3.5). Autant de nœuds d'entraînement sont créés, et légèrement plus de nœuds lors du contrôle de l'agent par l'IDSM, mais la dynamique reste la même. Le dernier cas avec  $k_d = 10000$  est intéressant car il provoque des zones où la variation de la commande motrice  $\boldsymbol{\mu}$  est nulle (dans les coins et au centre de l'espace sensorimoteur). Pour comprendre pourquoi, il faut expliquer notre calcul de  $d$ . Nous partons de l'équation 3.4 et multiplions numérateur et dénominateur par  $\exp(-k_d \|\mathbf{N}_p - \mathbf{x}_{SM}\|^2)$  pour éviter l'explosion numérique de  $\exp(k_d \|\mathbf{N}_p - \mathbf{x}_{SM}\|^2)$ . L'équation résultante est :

$$d(\mathbf{N}_p, \mathbf{x}_{SM}) = \frac{2 \exp(-k_d \|\mathbf{N}_p - \mathbf{x}_{SM}\|^2)}{1 + \exp(-k_d \|\mathbf{N}_p - \mathbf{x}_{SM}\|^2)} \quad (3.33)$$

Pour revenir sur notre cas avec  $k_d = 10000$ , dans les zones trop éloignées des nœuds, le terme  $\exp(-k_d \|\mathbf{N}_p - \mathbf{x}_{SM}\|^2)$  devient trop petit et est considéré comme nul (la limite est entre  $\exp(-745) = 5e - 324$  et  $\exp(-746) = 0$ ).  $d$  devenant nulle, la densité  $\phi_A$  devient nulle aussi, et suivant l'équation 3.11 la variation de  $\boldsymbol{\mu}$  est nulle à son tour. Nous revenons sur cette équation 3.11 et le choix lorsque  $\phi_A(\mathbf{x}_{SM}) < \epsilon$  en fin de cette section.

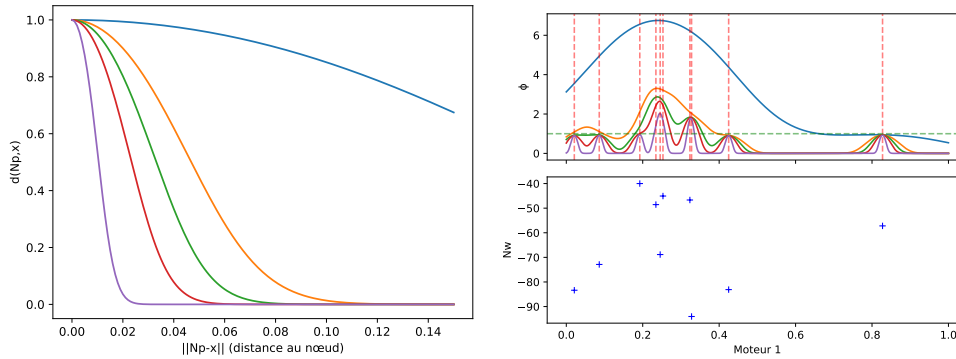
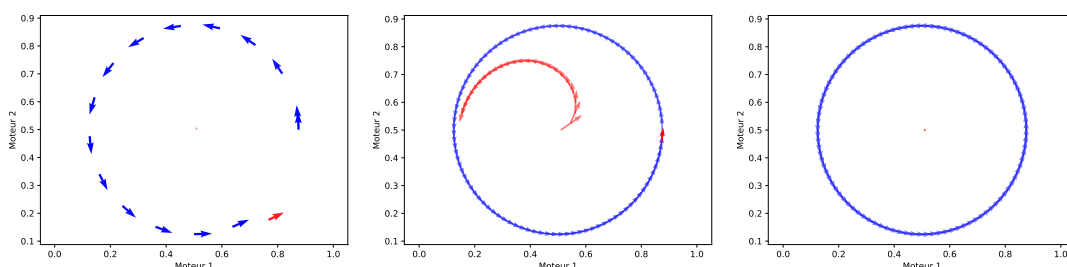


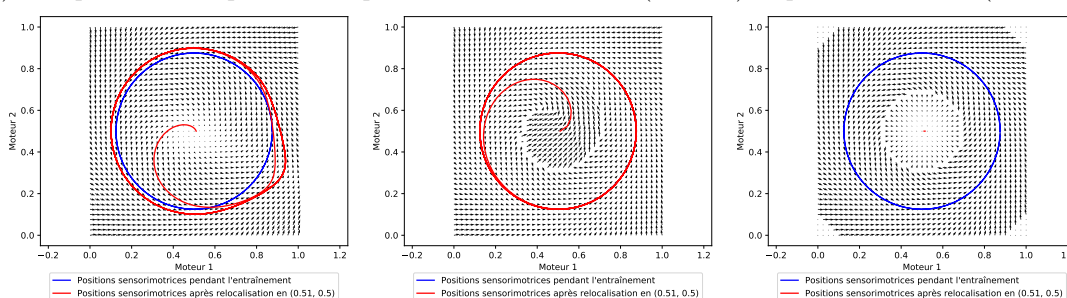
FIGURE 3.30 – Impact de  $k_d$  sur la fonction  $d$  et sur la densité  $\phi_N$ , avec  $k_d = 30$  en bleu, 500 en orange, 1000 en vert, 2000 en rouge et 10000 en violet. Plus le  $k_d$  augmente plus la zone d'influence du nœud rétrécit, et inversement. Sur le graphique de droite, les lignes rouges verticales pointillées indiquent la position des nœuds (dix nœuds placés aléatoirement avec des poids aléatoires dans  $[-100, 0]$ ). La ligne verte horizontale pointillée indique  $k_t$  le seuil de densité minimale pour la création de nouveaux nœuds.

### 3.4.5 Le paramètre de délai d'activation des nœuds $\tau$

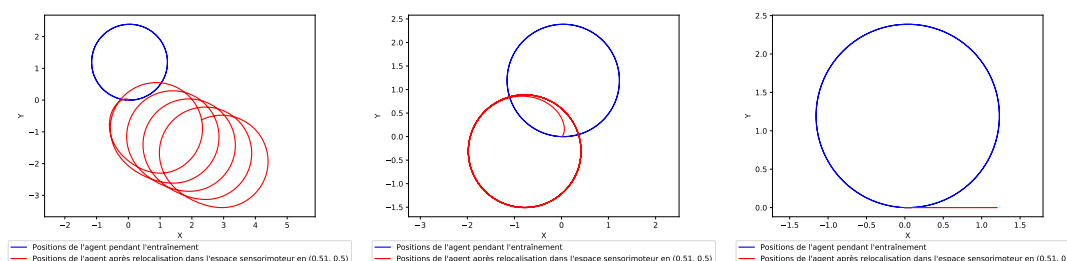
Le paramètre  $\tau$  est le délai nécessaire à un nœud pour devenir actif et influencer la mise à jour de la commande motrice  $\boldsymbol{\mu}$ . Dans le cas où il est nul, un nœud est actif dès sa création et du fait de la condition de la création de nœud (c'est-à-dire une faible densité de nœuds), ce nouveau nœud influence très fortement la commande motrice (voir figure 3.32 pour des exemples de  $\tau = 0$  avec le comportement cercle). Dans le cas d'un  $\tau$  non nul, l'impact de sa valeur sur la dynamique de l'espace sensorimoteur est complexe à évaluer car cela dépend du contexte lors de l'activation des nœuds. Cela dépend par exemple du nombre de nœuds déjà actifs, mais aussi de la position sensorimotrice de l'agent, de celle du nœud qui s'active et de celles des autres nœuds déjà actifs. Avant activation d'un nœud  $N$ , le comportement de l'agent sera toujours le même que si  $N$  n'existait pas, seuls les nœuds actifs contribuant à la variation de la commande motrice.



(a) Comparaison des poids créés pendant l'entraînement (en bleu) et pendant l'IDSM (en rouge).



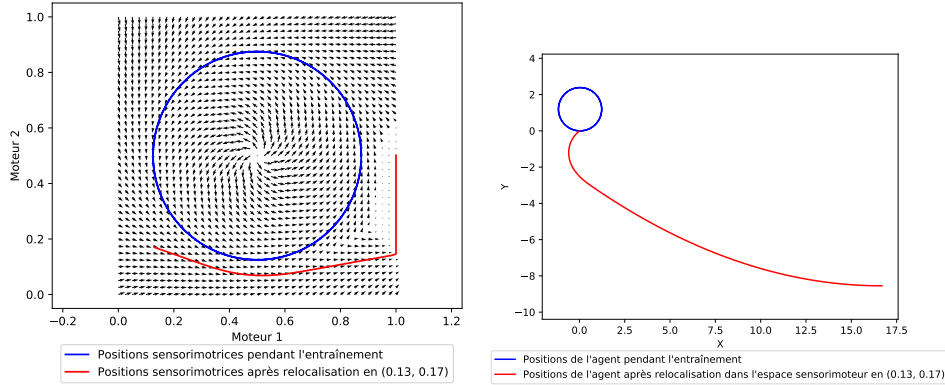
(b) Comparaison des trajectoires motrices pendant l'IDSM (en rouge) et des dynamiques en fin d'expérience.



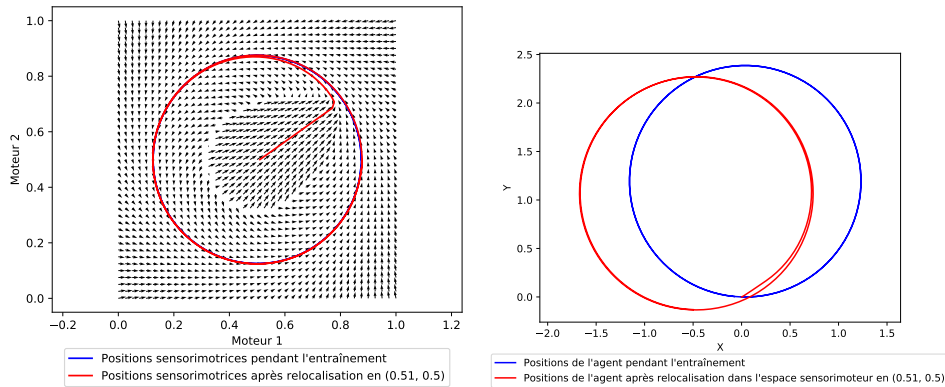
(c) Comparaison des trajectoires de l'agent (en rouge).

FIGURE 3.31 – Variation de  $k_d$  sur le comportement cercle avec une durée de 60 unités de temps.  $k_d = 30$  à gauche, 2000 au centre et 10000 à droite. Plus le  $k_d$  est faible, plus la zone d'influence des nœuds est étendue. À gauche, le déficit de nœuds dans le quartier en bas à droite du cercle renforce l'influence du dernier nœud bleu en bas, ce qui déforme la trajectoire motrice et par conséquent le comportement de l'agent (cercle glissant). Au centre, avec  $k_d = 2000$ , l'ensemble est similaire qu'avec  $k_d = 1000$  (voir figure 3.5). Dans le dernier cas, avec  $k_d = 10000$ , la zone d'influence des nœuds devient tellement faible qu'elle est considérée nulle au centre de l'espace sensorimoteur. Placer l'agent dans cette zone (ici en  $(0.51, 0.5)$ ) bloque sa commande motrice actuelle (variation de  $\mu$  nulle) et donc son comportement en cours : en l'occurrence, il se déplace légèrement vers la droite.

Si, lors de l'activation du nœud  $N$ , l'agent se trouve loin de lui et proche de nombreux autres nœuds actifs et influents, l'impact du nouveau nœud sera négligeable. À l'inverse, si le nœud le plus proche de l'agent est celui qui s'active, son impact sera important.



(a) Trajectoires de l'agent pour le comportement cercle avec  $\tau = 0$  en partant de (0.13, 0.17)



(b) Trajectoires de l'agent pour le comportement cercle avec  $\tau = 0$  en partant de (0.51, 0.5)

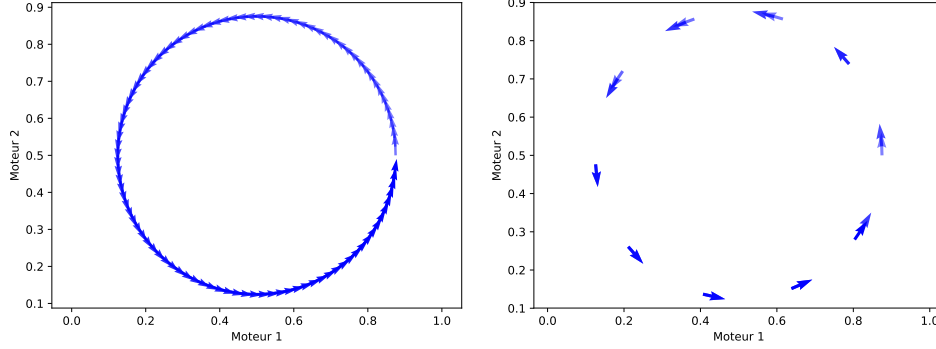
FIGURE 3.32 – Exemples du comportement cercle sans délai avant activation des nœuds ( $\tau = 0$ ). L'écart d'influence entre le nœud nouvellement créé, nécessairement le plus proche, peut être suffisant pour changer le comportement résultant de l'agent. C'est le cas dans l'exemple avec la relocalisation en extérieur du cercle (0.13, 0.17).

### 3.4.6 Délai minimum $\Delta t_{nodes}$ entre deux créations de nœud

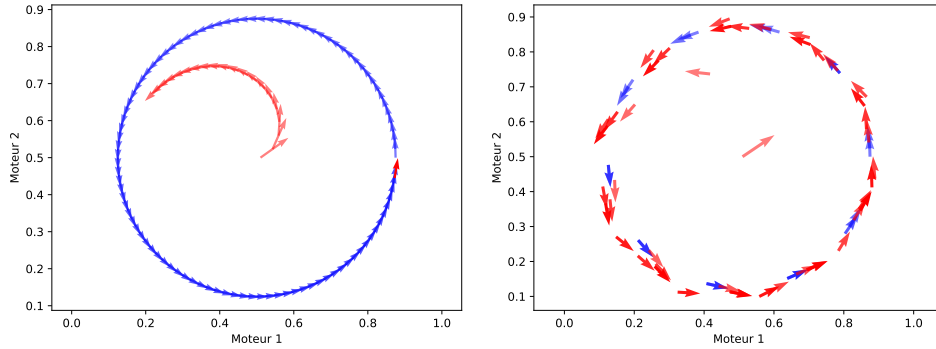
$\Delta t_{nodes}$  est l'écart de temps entre chaque tentative de création de nœud, donc entre chaque vérification que la densité des nœuds  $\phi_N$  est inférieure ou non au seuil  $k_t$ . Par défaut, nous fixons  $\Delta t_{nodes} = \Delta t = 0.1$ , donc la création de nœud se fait dès que possible. Les figures 3.33 et 3.34b montrent le résultat du comportement cercle lorsque  $\Delta t_{nodes}$  est augmenté à 1 pour ralentir la création de nœud à une fois par unité de temps. Contrairement à l'expérience initiale, nous laissons tourner l'IDSM pendant 60 unités de temps. Les trajectoires pendant l'entraînement sont identiques à celles "normales", la marche forcée ne dépendant pas des nœuds. Par contre, le nombre réduit de nœuds créés pendant l'entraînement impacte la trajectoire dans l'espace moteur de l'agent régit par l'IDSM (figure 3.34a). En effet, cette trajectoire se retrouve plus anguleuse. Le changement dans le comportement de l'agent (cercle glissant, voir figure 3.34b) provient du gain d'influence de chaque nœud par rapport à l'influence totale et de leur répartition plus chaotique. L'entraînement se termine avec moins de nœuds et une densité plus éparse (voir figure 3.33a).



Les nœuds créés après l'entraînement maintiennent durablement la dérive de trajectoire initiée par les nœuds centraux, en particulier celui proche du cercle d'entraînement vers  $[0.2, 0.65]$  (voir figure 3.33b).



(a) Ensembles de nœuds en fin d'entraînement du comportement cercle.

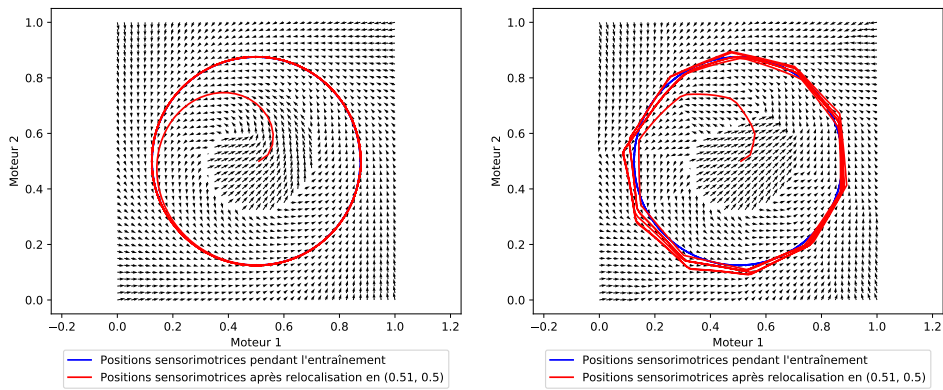


(b) Ensembles de nœuds en fin d'expérience.

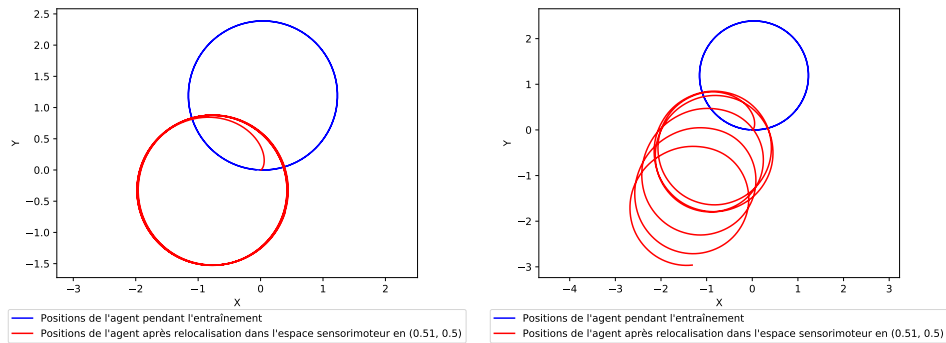
FIGURE 3.33 – Comparaison des nœuds créés entre à gauche avec une tentative de création de nœud dès que possible ( $\Delta t_{nodes} = \Delta t = 0.1$ ) et à droite qu'une fois par unité de temps ( $\Delta t_{nodes} = 1$ ). La faible couverture de l'entraînement dans le graphique de droite conduit à la fois à une influence plus forte des premiers nœuds créés pendant l'IDSM, et à la fois globalement à plus de nœuds créés sur le cercle. Ces derniers renforcent l'erreur au détriment de l'entraînement.

### 3.4.7 Seuil de saturation de nœuds $k_t$

Le paramètre  $k_t$  sert de seuil de saturation pour la densité  $\phi_{\mathcal{N}}$  en dessous duquel la création de nœud est possible. Si le seuil est bas, les nœuds sont peu nombreux et plus espacés. Un seuil élevé permet de créer plus de nœuds, donc de facilement enregistrer des perturbations dans la dynamique par rapport à celle d'entraînement, soit plus de plasticité dans le système. À l'inverse, un  $k_t$  faible renforce la rigidité du système car cela augmente indirectement la zone d'influence d'un nœud  $N$  en empêchant la création de nœuds concurrents aux alentours en fonction de son poids  $N_w$ . La figure 3.35 reprend l'exemple du cercle pour illustrer la différence d'impact sur l'IDSM entre faible et forte valeurs de  $k_t$ , avec  $k_t = 0.01$  à gauche et  $k_t = 10$  à droite. Contrairement à la section 3.4.6, la faible quantité de nœuds va être ici le cas où le comportement entraîné sera le plus stable. Peu de nœuds avec une faible saturation (la densité  $\phi_{\mathcal{N}}$  est inférieure à  $k_t$ , donc création possible de nouveaux nœuds) est une situation peu rigide, comme illustré sur la figure 3.33b. Sur le graphique de gauche de la figure 3.35a, le faible  $k_t$  donne une situation où il y a peu de nœuds mais une forte saturation, ce qui permet une forte rigidité.



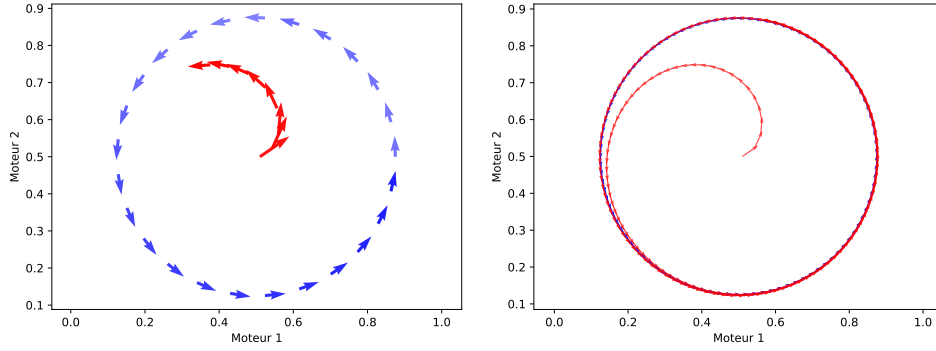
(a) Trajectoires dans l'espace moteur et dynamiques en fin d'expérience.



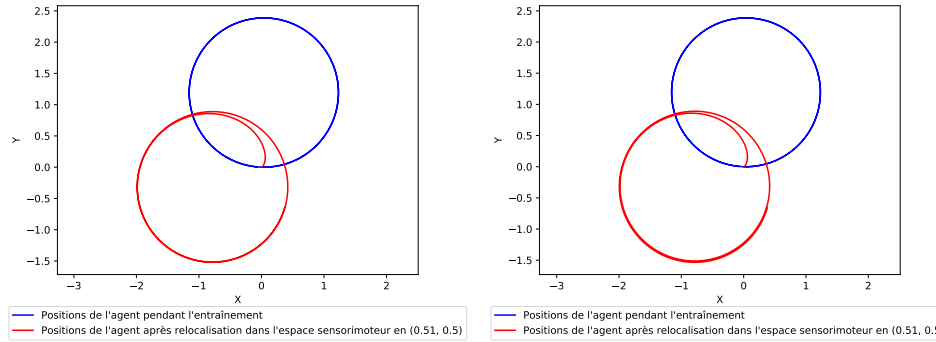
(b) Trajectoires de l'agent.

FIGURE 3.34 – Comparaison entre à gauche avec une tentative de création de nœud dès que possible ( $\Delta t_{nodes} = \Delta t = 0.1$ ) et à droite qu'une fois par unité de temps ( $\Delta t_{nodes} = 1$ ). La trajectoire motrice est moins stable, surtout dans le quart en bas à gauche, ce qui entraîne un comportement de cercle glissant.

En fin d'expérience, aucun nœud n'a été créé sur le cercle de nœuds d'entraînement pendant la phase où l'IDSM contrôle l'agent, l'agent se comportant comme pendant l'entraînement. Avec un fort  $k_t$ , il y a de nombreux nœuds, mais une faible saturation, il faudrait un entraînement plus long pour saturer l'espace. En conséquence, la situation est plastique, de nombreux nœuds sont créés pendant l'IDSM, provoquant une trajectoire motrice légèrement plus large que celle d'entraînement.



(a) Comparaison des nœuds créés, en bleu pendant l'entraînement et en rouge pendant le contrôle par l'IDSM.



(b) Comparaison des trajectoires des agents.

FIGURE 3.35 – Impact du paramètre  $k_t$ , avec  $k_t = 0.01$  à gauche et  $k_t = 10$  à droite. Un faible  $k_t$  veut dire une saturation plus rapide, donc peu de création de nœuds et un système plus rigide. Un fort  $k_t$  à l'inverse rend la saturation en nœuds plus lente à atteindre et un système plus plastique. Sur le second graphique de la figure 3.35b, le cercle rouge s'épaissit car l'influence des nœuds du centre est prise en compte lors de la création des nouveaux nœuds du cercle, poussant la trajectoire motrice plus loin.

### 3.4.8 Commande motrice $\mu$

La position de l'agent dans l'espace sensorimoteur,  $\mathbf{x}_{SM}$ , est nécessairement maintenue dans cet espace normalisé, espace qui représente l'étendue des capacités sensorielles et motrices de l'agent. Le cas de la commande motrice  $\mu$  n'est pas si simple à trancher : doit-elle être maintenue dans  $[0, 1]$  ou non ? Dans nos premières expériences, nous avons fait le choix que non, et nous avons  $\mu \in [-\infty; +\infty]$ . La figure 3.36 illustre à l'aide d'un cas simple l'impact sur la trajectoire de l'agent des deux possibilités. La différence peut sembler négligeable, mais ne pas borner  $\mu$  à l'espace moteur revient à l'utiliser comme accumulateur lorsqu'au moins un moteur est à une de ses limites et rajoute de l'inertie. Plus la dynamique pousse l'agent en dehors du cadre, et plus

difficilement le ou les moteurs concernés vont pouvoir être modifiés. De plus, même sur un cas simple comme celui illustré ici, ce changement de trajectoire entraîne la création de nouveaux nœuds, donc une modification de la dynamique globale.

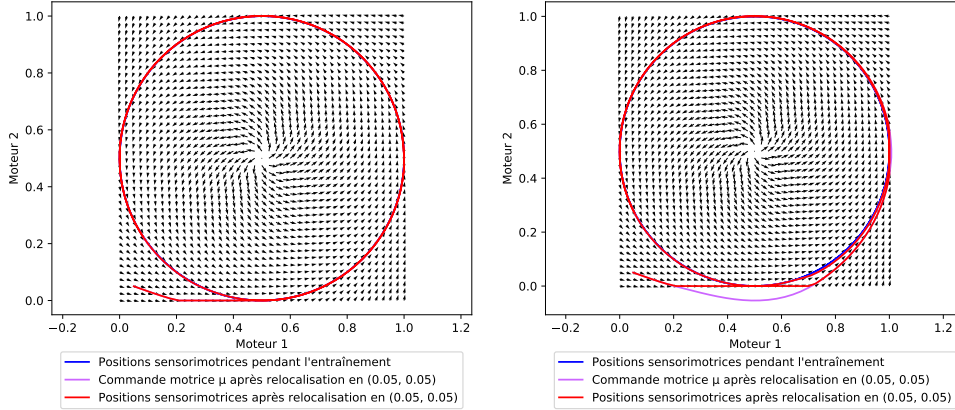


FIGURE 3.36 – Dynamiques en fin d'expérience, à gauche en limitant la commande motrice  $\mu \in [0, 1]$ , à droite avec  $\mu \in [-\infty; +\infty]$ . La valeur de  $\mu$  est tracée en mauve, elle n'est pas visible à gauche car strictement égale à la position sensorimotrice de l'agent (en rouge). Sur le graphique de gauche, l'agent "récupère" la trajectoire d'entraînement dès que possible (la dynamique étant équilibrée, dès  $(m_1, m_2) = (0.5, 0)$ ). À droite, il ne superpose la trajectoire d'entraînement qu'après  $(m_1, m_2) = (1, 0.5)$  et un décalage permanent reste visible entre trajectoires en bas à droite du cercle à cause de l'influence des nouveaux nœuds.

### 3.4.9 Calcul de la variation de la commande motrice $\dot{\mu}(t)$

Dans les cas où l'agent est loin des nœuds actifs, ce qui peut arriver assez vite en fonction du nombre de dimensions par exemple, le calcul de la variation de la commande motrice  $\mu$  est problématique (voir équation 3.11). En effet, le dénominateur impose que la densité des nœuds actifs  $\phi_{\mathcal{A}}(\mathbf{x}_{SM})$  soit non nulle. En théorie, avec des réels,  $\phi_{\mathcal{A}}(\mathbf{x}_{SM})$  ne peut jamais être nul, mais la limite de précision des calculs par ordinateurs fait que l'on peut obtenir un dénominateur nul. Le choix que nous avons fait est de filtrer le cas problématique avec un paramètre  $\epsilon$ , et renvoyer une variation nulle de la commande motrice  $\mu$  lorsque  $\phi_{\mathcal{A}}(\mathbf{x}_{SM}) \leq \epsilon$ .

Cependant ce choix n'est pas sans conséquences, car en fonction de la valeur  $\epsilon$  cela peut créer des zones plus ou moins grandes sans variations de  $\mu$  où elles n'auraient pas lieu d'être. Remplacer le dénominateur par  $\epsilon$  plutôt que de renvoyer 0 résulte en une chute légèrement moins brutale mais ne change pas fondamentalement le problème.

Une autre solution, non testée, serait de fixer une valeur minimale à  $d(\mathbf{N}_p, \mathbf{x}_{SM})$ . La contrepartie serait qu'au delà d'une certaine distance euclidienne entre un nœud et l'agent, le nœud serait artificiellement rapproché de l'agent. Sans analyses, il est compliqué de savoir si l'impact sur une dynamique complexe avec de nombreux nœuds serait significatif ou non.

## 3.5 Motivations

De nombreuses caractéristiques de l'IDSM nous intéressent. Il permet un apprentissage en temps continu de trajectoires sensorimotrices effectuées par l'agent en vu de lui faire répéter ce

qu'il a déjà fait. Ces trajectoires peuvent être interprétées directement comme le comportement de l'agent, elles sont mémorisées par des structures localement disposées dans l'espace sensorimoteur et dont l'interaction crée une dynamique incitant l'agent à reproduire le même comportement. Dans une certaine mesure, l'IDSM permet aussi de faire émerger des trajectoires sensorimotrices inexplorées via cette dynamique de l'espace sensorimoteur, ce qui se traduit en une émergence de comportements non observés auparavant. Par exemple, comme illustré par la dernière expérience sur la combinaison de comportements utilisant un senseur (voir figures 3.23 et 3.24), un agent peu apprendre un certain comportement dans un certain contexte sensoriel, et un autre comportement dans un autre contexte sensoriel. À partir de là, s'il se est dans un contexte sensoriel intermédiaire, il se retrouve dans une dynamique résultante de l'influence des nœuds des deux comportements et cette dynamique peut être source d'un nouveau comportement. L'avantage de cet espace sensorimoteur est qu'il contient l'intégralité des trajectoires sensorimotrices que *pourrait* suivre l'agent, qui sont à la fois les comportements qu'il a déjà explorés et ceux qu'il ne connaît pas encore. La question est alors de comment faire pour aider l'agent à aller davantage du côté de l'exploration de ses possibilités sensorimotrices plutôt que de rester du côté de l'exploitation de ce qu'il connaît déjà, ce qui est favorisé par l'IDSM par défaut.

Si nous nous replaçons sur les modifications de comportements qui nous intéressent (modulation, exploration et acquisition), le modèle IDSM permet déjà de lui-même d'y répondre en partie. Concernant la modulation, où l'agent adapte son comportement au contexte, ce n'est pas le cas. L'agent ne peut pas de lui-même adapter son comportement. Cependant, le modèle IDSM propose déjà un cadre dans lequel une modification de contexte permet une modification de comportement. L'expérience des des comportements combinés avec senseur (voir les figures 3.23, 3.24 et 3.25) illustre les possibilités de modulations de comportements dans un cadre simple, où changer la valeur du senseur emmène l'agent dans une dynamique différente. Pour l'exploration autour des comportements prédéfinis ou acquis, l'IDSM n'inclut pas d'aléatoire mais la dynamique résultante des nœuds permet déjà, dans une certaine mesure, à l'agent de reproduire des trajectoires sensorimotrices pas toujours parfaitement identiques. Les nœuds mémorisent individuellement la vitesse de l'agent à leur création, mais la trajectoire sensorimotrice de l'agent est définie par la dynamique résultant de l'ensemble de ces nœuds. Malgré la composante attraction qui pousse l'agent vers les nœuds, la dynamique ne résulte pas nécessairement en une trajectoire déjà expérimentée par l'agent, surtout lorsque le comportement de l'agent impacte indirectement sa perception comme dans l'expérience avec une lampe fixe (voir section 3.2.2). Cette légère instabilité est en plus accentuée par la création de nouveaux nœuds et leur délai d'activation ainsi que par la modification des poids de ces nœuds dans le temps en fonction de leur distance à la position sensorimotrice de l'agent. Outre des modifications de trajectoires sensorimotrices à la marge, résultantes en des comportements similaires, ce jeu de dynamique peut aller jusqu'à faire émerger des comportements non observés par l'agent auparavant, comme par exemple les cas des figures 3.12c et 3.13, où l'agent fait des cercles sans avoir appris ce comportement au préalable. Enfin, sur l'acquisition de comportements, l'IDSM permet de mémoriser à l'aide des nœuds tout comportements impliquant une zone de l'espace sensorimoteur ayant été peu ou pas exploitée. La condition pour cette mémorisation est cependant compliquée à prévoir, l'exemple de combinaison de comportement montrant qu'un comportement appris peut "facilement" en supplanter un autre s'il a été mémorisé ou ré-exprimé par l'agent peu de temps auparavant (voir section 3.3.1, figures 3.18 et 3.19).

Dans la suite de ce manuscrit, nous présentons nos principales contributions qui sont les modèles permettant de doter un agent basé sur l'IDSM des différents apprentissages non-associatifs

que sont l'habituation, la sensibilisation spécifique au site et le pseudo-conditionnement. Ces apprentissages vont principalement permettre, dans les configurations simples que nous avons testées, d'ajouter de la modulation de comportements au contexte environnemental (stimulation continue ou répétée).

# IDSMS avec apprentissages non-associatifs

Nous nous intéressons maintenant à intégrer les différents apprentissages non-associatifs (habituations, sensibilisation spécifique au site et pseudo-conditionnement) à un agent contrôlé par l'IDSMS. Nous souhaitons ainsi permettre à l'agent d'explorer différemment son espace sensorimoteur, en étant guidé par ces apprentissages pour atteindre des zones inexplorées de son espace sensorimoteur ou pour modifier les trajectoires déjà existantes.

Pour simplifier les analyses, nous désactivons la création de nœuds et la mise à jour des poids des nœuds de l'IDSMS. Cependant, pour chacun des apprentissages, nous discutons ce qui se passerait avec l'IDSMS complet, soit en se basant sur des travaux préliminaires, soit en détaillant nos hypothèses à travers notre compréhension de nos modèles proposés.

Dans une première section, nous traitons de l'habituations et proposons trois modèles permettant à l'agent d'inhiber sa réponse à un stimulus. Une seconde section concerne la sensibilisation spécifique au site et adapte les trois modèles pour cet apprentissage. Ensuite, une troisième section propose un quatrième modèle focalisé sur le pseudo-conditionnement. Enfin, nous concluons ce chapitre par un bilan de nos différents modèles.

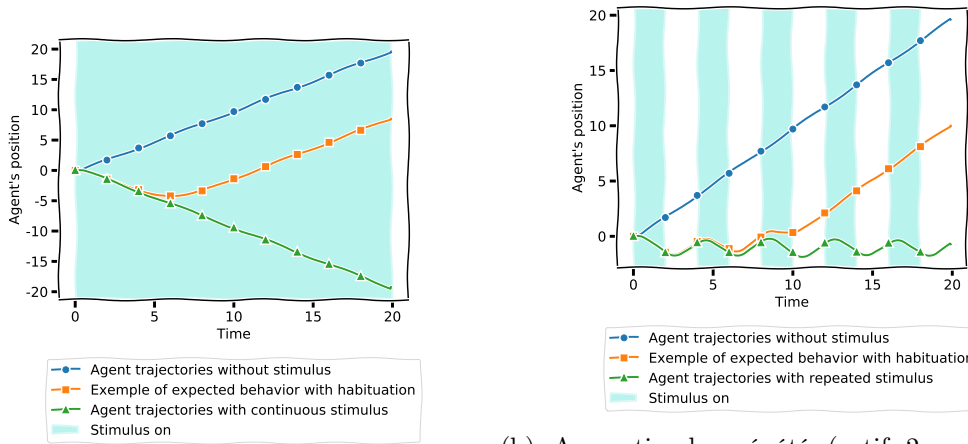
## 4.1 Habituations

Afin d'intégrer l'habituations telle que définie dans la section 2.1.2 dans l'IDSMS, nous proposons trois modèles. Dans le premier, une mémoire sensorielle inhibe la perception du stimulus. Dans les deux autres, l'habituations se produit par l'affaiblissement de l'influence des nœuds régulièrement sollicités, soit à travers une composante énergétique (que l'on a ajouté à chaque nœud), soit par la modification du poids des nœuds.

Le modèle avec mémoire sensorielle a été pensé pour des stimuli continus (par exemple une pile de papiers à ranger qui traîne longuement sur un bureau que l'on finit par ne plus voir) là où les modèles avec pénalité sur les nœuds ont été conçus pour gérer des stimuli brefs et répétés (par exemple un néon défectueux).

Nous illustrons le comportement que nous attendons de nos modèles à l'aide d'une expérience de

### Illustration de l'habituation et de la récupération spontanée



(a) Avec stimulus continu

(b) Avec stimulus répété (actif 2 unités de temps, inactif 2 u.t.)

FIGURE 4.1 – Illustration des comportements attendus avec ou sans habituation. Hors stimulation, le comportement prédéfini de l'agent est de monter. Lorsque le stimulus est perçu, son comportement prédéfini est de descendre. Avec l'habituation, nous nous attendons à ce que l'agent apprenne progressivement à se comporter comme hors stimulation lorsque la stimulation est continue ou répétée (c'est-à-dire monter à la place de descendre).

pensée, illustrée sur la figure 4.1. L'objectif est d'obtenir une diminution de la réponse lorsqu'un stimulus est actif et une récupération partielle de la réponse lorsque le stimulus est absent. Nous utilisons une configuration avec un agent simple, avec seulement un moteur et un senseur. Il existe un stimulus dans l'environnement qui peut être activé ou désactivé. Lorsque le stimulus est actif, l'agent le perçoit avec une valeur constante quelle que soit la position de l'agent dans l'environnement. L'agent peut se déplacer le long d'une seule dimension, vers le haut ou vers le bas. Son comportement réflexe est de monter en l'absence de stimulus et de descendre lorsque l'agent perçoit le stimulus. Dans un tel contexte, une habituation réussie devrait se traduire par l'inhibition de la réponse (mouvement descendant) après une présentation continue ou répétée du stimulus : l'agent se met alors à monter à la place de descendre.

Dans la figure 4.1, nous considérons deux configurations et, pour chacune, nous montrons comment l'agent est censé se comporter en l'absence de stimulus (toujours la trajectoire du haut), en présence d'un stimulus en étant incapable d'habituation (toujours la trajectoire du bas) et en présence d'un stimulus en étant capable d'habituation. Lorsqu'un stimulus est présenté, il peut être continu ou répété et, pour faciliter la lecture, les périodes où le stimulus est actif sont représentées par un fond coloré.

Dans la première configuration (figure 4.1a), le stimulus est activé en continu. Dans la deuxième configuration (figure 4.1b), le stimulus est activé de manière répétée pendant la moitié du temps. Les trajectoires du haut (sans stimulus) et du bas (sans habituation) servent de référence. Un agent capable d'habituation devrait commencer par suivre la trajectoire du bas, puis s'en détacher progressivement et tendre vers la trajectoire du haut, pour finalement être parallèle à celle-ci une fois le stimulus totalement ignoré.



Notons que ces graphiques ne sont pas les résultats de simulation mais sont "dessinés à la main" pour illustrer les propriétés de l'habituation.

### 4.1.1 Modèles

#### Modèle avec mémoire sensorielle

L'habituation est définie comme une réduction progressive de l'un ou plusieurs paramètres d'une réponse à un stimulus répété. Le premier modèle que nous proposons se base sur une mémoire sensorielle  $\delta$  intégrant la valeur du senseur et inhibant en proportion la perception de l'agent du stimulus capté par ce senseur. Cette baisse de perception provoque un changement dans l'expression de la réponse car elle implique un changement dans la composante sensorielle dans la position sensorimotrice  $\mathbf{x}_{SM}$  de l'agent. La mémoire sensorielle  $\delta$  garde trace de l'évolution du stimulus perçu. La présentation répétée d'un stimulus va faire croître progressivement cette mémoire et l'on peut utiliser l'écart entre la mémoire et la valeur du stimulus pour atténuer la perception qu'a l'agent du stimulus. Avec ce modèle, la valeur sensorielle considérée par l'IDSM est la valeur du senseur *habitué*  $s_h$ , qui est modulée par le facteur d'apprentissage  $f_a$  (voir équations 4.2-4.4). Ce facteur est fonction de la différence entre la valeur sensorielle brute  $s_b$  et la moyenne de celle-ci sur une fenêtre de temps récente (la mémoire sensorielle  $\delta$ ). Plus la mémoire sensorielle  $\delta$  est proche de la valeur sensorielle brute  $s_b$ , plus la réponse motrice est inhibée, sans aucune modification sur les nœuds de l'IDSM.

Si le stimulus cesse d'être présenté, la valeur du senseur est 0, la mémoire sensorielle  $\delta$  va alors redescendre vers 0 et la réponse motrice redevient progressivement la réponse initiale (récupération spontanée).

La mémoire sensorielle  $\delta$  est calculée comme un filtre passe-bas sur la valeur sensorielle brute  $s_b$  :

$$\frac{d\delta}{dt}(\mathbf{x}_E, t) = \beta(s_b(\mathbf{x}_E) - \delta(\mathbf{x}_E, t)) \quad (4.1)$$

avec  $\beta = 0.25$ . Ce paramètre  $\beta$  (compris entre 0 et 1) règle la sensibilité du modèle aux variations du stimulus : plus il est élevé, plus l'est la sensibilité et plus rapidement le modèle atteint un niveau asymptotique.

La valeur du senseur *habitué*  $s_h$ , utilisée en tant que composante sensorielle de la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$ , est calculée à partir de la valeur sensorielle brute  $s_b$  et de la mémoire sensorielle  $\delta$  :

$$s_h(\mathbf{x}_E, t) = f_a(\mathbf{x}_E, t)s_b(\mathbf{x}_E) \quad (4.2)$$

$$f_a(\mathbf{x}_E, t) = \begin{cases} \frac{1}{1 + e^a} & \text{si } s_b(\mathbf{x}_E) > \epsilon \\ 0 & \text{sinon} \end{cases} \quad (4.3)$$

$$a = -10 \left( \frac{\|\delta(\mathbf{x}_E, t) - s_b(\mathbf{x}_E)\|_2}{s_b(\mathbf{x}_E)} - \frac{1}{2} \right) \quad (4.4)$$

La figure 4.2 illustre le comportement du facteur d'apprentissage  $f_a$ . Le stimulus est fortement perçu lorsque la valeur sensorielle brute  $s_b$  est éloignée de la mémoire sensorielle (on a  $\|\delta - s_b\|$  proche de  $s_b$ ) et il est fortement inhibé si la valeur sensorielle brute est similaire à la mémoire sensorielle ( $\delta \approx s_b$ ).

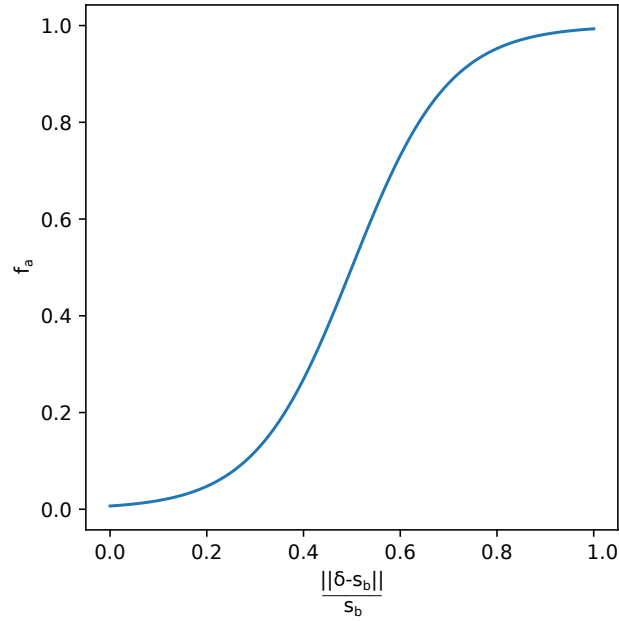


FIGURE 4.2 – Le facteur d'apprentissage  $f_a$  comme fonction de la différence entre la valeur sensorielle brute  $s_b(\mathbf{x}_E)$  et la mémoire sensorielle  $\delta$ .

### Modèles avec pénalité sur les nœuds

Alors que le modèle précédent affecte globalement la perception sensorielle, il est également possible d'implémenter l'habituation à travers des mécanismes locaux, au niveau des nœuds de l'IDSM. Chaque nœud peut être affecté localement pour produire l'habituation en inhibant localement son influence sur la réponse motrice.

Les deux modèles suivants sont basés sur le concept de surprise dans l'espace sensorimoteur. L'idée est que chaque fois que l'agent est "surpris" par sa position réelle dans l'espace sensorimoteur, les nœuds environnants sont pénalisés. Plus un nœud est impacté, plus son influence sur la dynamique IDSM est diminuée. Notre intuition est que les nœuds préexistants encodent le comportement de réponse et que, lorsqu'il est "surpris", l'agent diminue l'influence de ce comportement de réponse, conduisant ainsi à l'habituation. Pour permettre une récupération spontanée, l'effet de la pénalité sur un nœud impacté par la surprise s'estompe progressivement dans le temps.

La surprise  $sur$  est définie comme une erreur entre la prédiction de la position sensorimotrice attendue  $\hat{\mathbf{x}}_{SM}$  et la position sensorimotrice observée  $\mathbf{x}_{SM}$ .

$$sur(t) = \|\mathbf{x}_{SM}(t) - \hat{\mathbf{x}}_{SM}(t)\|_2 \quad (4.5)$$

$$\hat{\mathbf{x}}_{SM}(t) = \mathbf{x}_{SM}(t - \Delta t) + \begin{bmatrix} \dot{\mathbf{x}}(t - \Delta t) \\ 0 \end{bmatrix} \Delta t \quad (4.6)$$

Si l'environnement n'impacte pas la commande motrice (ce qui est le cas dans nos expériences) alors avec cette définition de la prédiction  $\hat{\mathbf{x}}_{SM}$ , nous pouvons observer que  $[\mathbf{x}_{SM}(t)]^\mu = [\hat{\mathbf{x}}_{SM}(t)]^\mu$ . Cela entraîne que la surprise *sur* vaut toujours  $[\mathbf{x}_{SM}(t)]^\sigma - [\hat{\mathbf{x}}_{SM}(t)]^\sigma$ , avec  $[\mathbf{x}_{SM}(t)]^\sigma$  la projection dans l'espace senseur. Par conséquent, dans nos expériences, la surprise est toujours sensorielle : c'est un changement dans la perception des stimuli. De plus, cela implique que dans ce cas la surprise est toujours dans  $[0, \sqrt{K^\sigma}]$ , avec  $K^\sigma$  le nombre de dimensions sensorielles (soit le nombre de senseurs de l'agent).

**Impact sur l'énergie des nœuds** Nous élaborons un premier mécanisme local en ajoutant une composante énergétique  $N_e \in [0, 1]$  à chaque nœud  $N$ . L'influence d'un nœud sur la dynamique de l'IDSM dépend alors de sa valeur énergétique. Plus l'énergie est faible, plus l'influence est faible. La surprise *sur* diminue l'énergie des nœuds qui est récupérée progressivement dans le temps.

En pratique, les facteurs de vitesse et d'attraction sont modifiés afin de prendre en compte l'énergie des nœuds :

$$\mathbf{V}(\mathbf{x}_{SM}) = \sum_{\mathcal{A}} \xi_N(\mathbf{x}_{SM}) \cdot N_e \cdot [\mathbf{N}_v]^\mu \quad (4.7)$$

$$\mathbf{A}(\mathbf{x}_{SM}) = \sum_{\mathcal{A}} \xi_N(\mathbf{x}_{SM}) \cdot N_e \cdot [\mathbf{\Gamma}(\mathbf{N}_p - \mathbf{x}_{SM}, \mathbf{N}_v)]^\mu \quad (4.8)$$

$$\xi_N(\mathbf{x}_{SM}) = \omega(N_w) \cdot d(\mathbf{N}_p, \mathbf{x}_{SM}) \quad (4.9)$$

L'énergie consommée  $c_e$  d'un nœud  $N$  dépend de la surprise *sur* et de la proximité entre  $N$  et la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$  :

$$c_e(N, t) = ((1 + \text{sur}(t))^{\beta_e} - 1) \cdot \xi_N(\mathbf{x}_{SM}) \quad (4.10)$$

$$N_e(t + \Delta t) = N_e(t) + \gamma_e \Delta t - c_e(N, t) \quad (4.11)$$

où  $\beta_e = 0.5$  régule la consommation d'énergie et  $\gamma_e = 0.05$  régule la récupération d'énergie.

**Impact sur le poids des nœuds** Nous élaborons un second mécanisme local en modulant le poids des nœuds. La valeur du poids diminue en cas de surprise *sur* et se rétablit progressivement au fil du temps. Ce modèle est plus conforme à l'IDSM original.

$$c_w(N, t) = \beta_w \cdot \text{sur}(t) \cdot \xi_N(\mathbf{x}_{SM}) \quad (4.12)$$

$$N_w(t + \Delta t) = N_w(t) + \gamma_w \Delta t - c_w(N, t) \quad (4.13)$$

avec  $\beta_w = 500$  régulant la pénalité de poids et  $\gamma_w = 75$  pour la récupération spontanée. Les poids des nœuds sont conservés dans  $] -\infty, 0]$ .

Ainsi, pour un nœud donné  $N$ , son poids  $N_w$  est impacté par l'apprentissage lorsque la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$  est proche de la position du nœud  $\mathbf{N}_p$  et que la surprise n'est pas 0. Une telle surprise ne se produit que pour les événements liés à l'activation ou à la désactivation du stimulus.

$\Delta t$	$k_t$	$k_d$	$k_\omega$	$\tau$	$\mu$
0.1	1	30	0.025	10	$\in [-\infty; +\infty]$

TABLE 4.1 – Valeurs des paramètres et options utilisées pour les résultats sur l’habituation (section 4.1.2)

### 4.1.2 Expériences et résultats

Dans cette section, nous présentons l’instanciation du modèle IDSM pour les expériences sur l’habituation sur un agent avec un moteur et un senseur pouvant se déplacer sur un axe. Ensuite, nous expliquons la configuration expérimentale utilisée pour évaluer les effets de l’habituation et de la récupération spontanée provoqués par les modèles présentés précédemment, et nous soulignons certaines caractéristiques de ces modèles. Ensuite, nous explorons un peu plus l’effet de l’intervalle interstimulus (caractéristique 4, rappelée avant présentation des résultats associés) et enfin les résultats de la combinaison des modèles.

#### Instanciation du modèle

Pour nos expériences, nous construisons un IDSM avec deux nœuds comme indiqué sur la figure 4.3. Les deux nœuds sont définis par :

- $N^1 = \langle \mathbf{p} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, w = 0 \rangle$ ,
- $N^2 = \langle \mathbf{p} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, w = 0 \rangle$ .

Ces deux nœuds conduisent à la dynamique motrice illustrée par les flèches noires du graphique et produisent le comportement attendu : si l’agent perçoit le stimulus (la valeur sensorielle brute est  $s = 1$ ), sa vitesse diminue (la composante motrice tend vers 0) car  $N^1$  est le nœud le plus influent ; si l’agent ne perçoit pas le stimulus (la réponse sensorielle est  $s = 0$ ), sa vitesse augmente (la composante motrice tend vers 1) car  $N^2$  est le nœud le plus influent. La composante motrice  $[\mathbf{x}_{SM}]^\mu$  est le moteur normalisé dans l’espace sensorimoteur : l’agent descend plus ou moins vite lorsque  $[\mathbf{x}_{SM}]^\mu$  est dans  $[0; 0.5[$ , l’agent est à l’arrêt lorsque  $[\mathbf{x}_{SM}]^\mu$  vaut 0.5 et l’agent monte plus ou moins vite lorsque  $[\mathbf{x}_{SM}]^\mu$  est dans  $]0.5, 1]$ .

#### Configuration expérimentale pour l’habituation et la récupération spontanée

Afin de mesurer si nos modèles sont capables d’habituation et de récupération spontanée, nous utilisons le protocole suivant.

L’expérience est divisée en différentes phases : des phases de stimulation (avec un stimulus continu ou répété) et des phases de calme (sans aucun stimulus). Il y a d’abord deux phases de stimulation pour évaluer l’habituation, puis une phase de calme suivie d’une phase de stimulation pour évaluer la récupération spontanée. Chaque phase dure vingt unités de temps. Pour limiter les différences entre les conditions expérimentales, au début de chaque phase, l’agent est réinitialisé à une position de départ à vitesse nulle ( $\mathbf{x}_E = \dot{\mathbf{x}}_E = \ddot{\mathbf{x}}_E = 0$ ) et à un état sensorimoteur neutre  $\mathbf{x}_{SM} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$ . Ici, nous avons choisi de forcer la valeur du senseur à 0 pour avoir des conditions initiales identiques entre les phases, en particulier pour comparer la première et la quatrième phase.

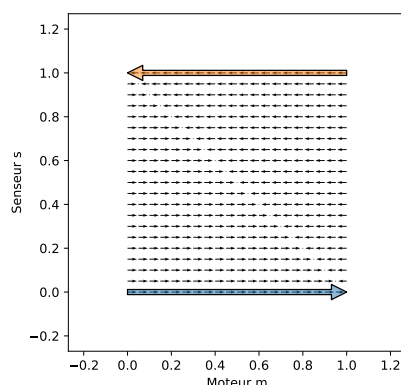


FIGURE 4.3 – IDSM avec deux nœuds utilisé dans les expériences sur l’habituation. Les flèches colorées représentent les nœuds, les flèches noires indiquent la dynamique motrice : la composante motrice  $[\mathbf{x}_{SM}]^\mu$  tend vers 0 lorsque le stimulus est perçu ou vers 1 lorsque le stimulus n’est pas perçu. Les comportements liés au moteur sont que l’agent descend plus ou moins vite lorsque  $[\mathbf{x}_{SM}]^\mu$  est dans  $[0; 0.5[$ , l’agent est à l’arrêt lorsque  $[\mathbf{x}_{SM}]^\mu$  vaut 0.5 et l’agent monte plus ou moins vite lorsque  $[\mathbf{x}_{SM}]^\mu$  est dans  $]0.5, 1]$ .

Dans une expérience donnée, la phase de stimulation suit l’un des quatre schémas ci-dessous. Notons que pour une expérience, toutes les phases de stimulation suivent le même schéma.

- (a) : stimulus continu ;
- (b) : répétitions de 3 unités de temps avec stimulus / 1 unité de temps sans ;
- (c) : répétitions de 2 unités de temps avec / 2 unités de temps sans ;
- (d) : répétitions de 1 unité de temps avec / 3 unités de temps sans.

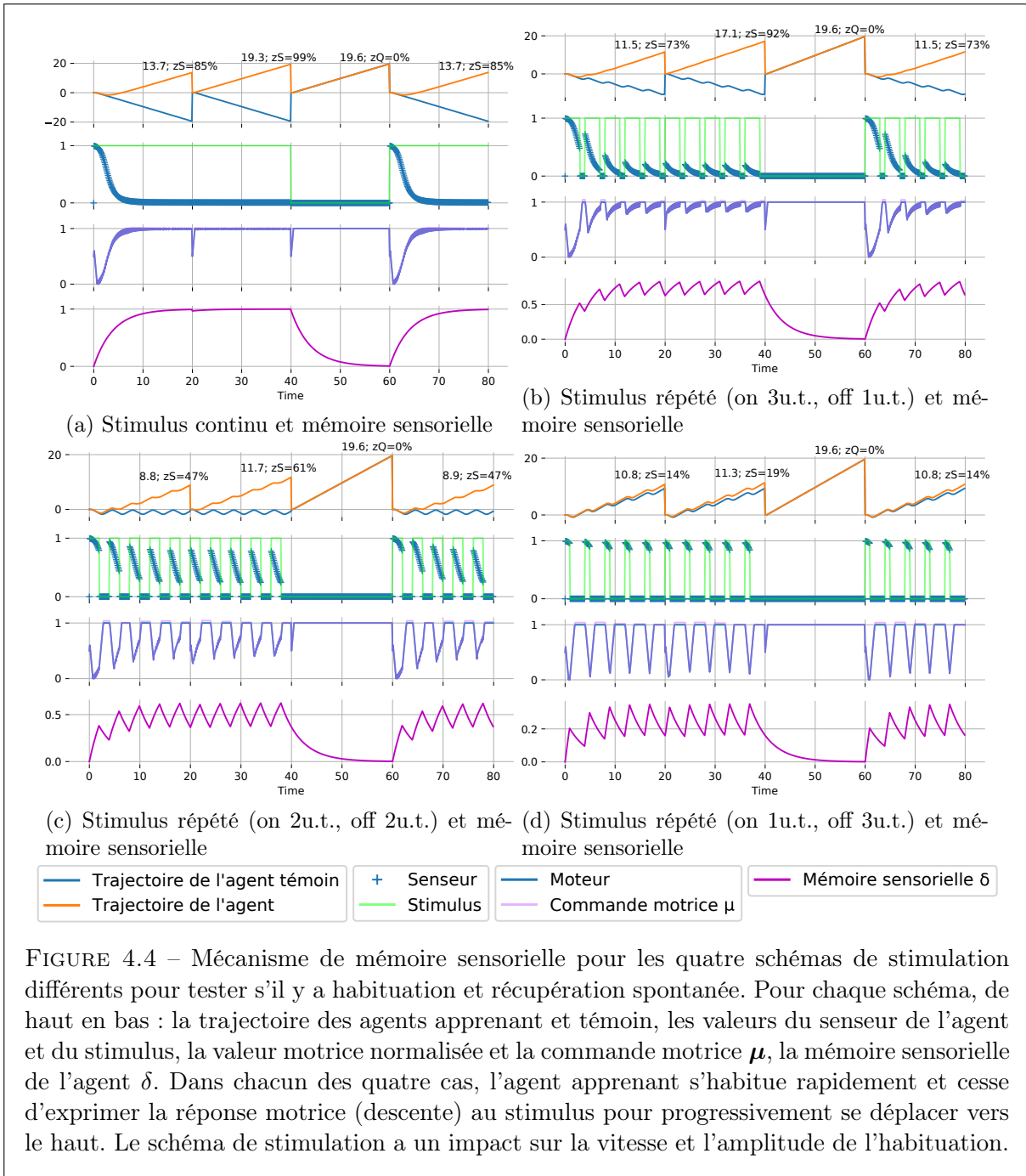
### Résultats pour l’habituation et la récupération spontanée

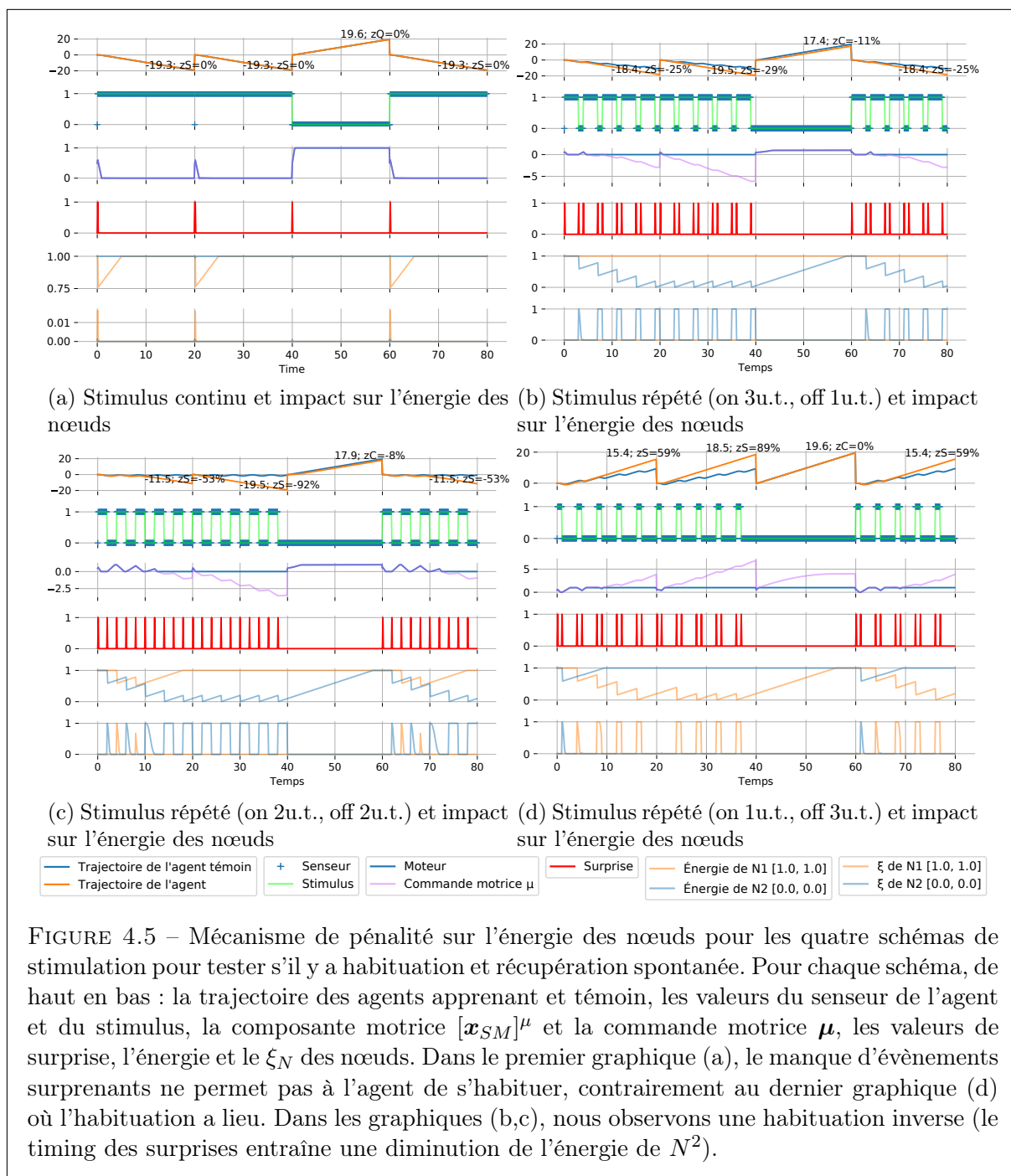
L’habituation et la récupération spontanée sont évaluées en comparant le comportement d’un agent apprenant (utilisant un de nos modèles) avec celui d’un agent témoin (sans mécanisme d’apprentissage). Les résultats pour chaque modèle sont donnés sur les figures 4.4-4.6.

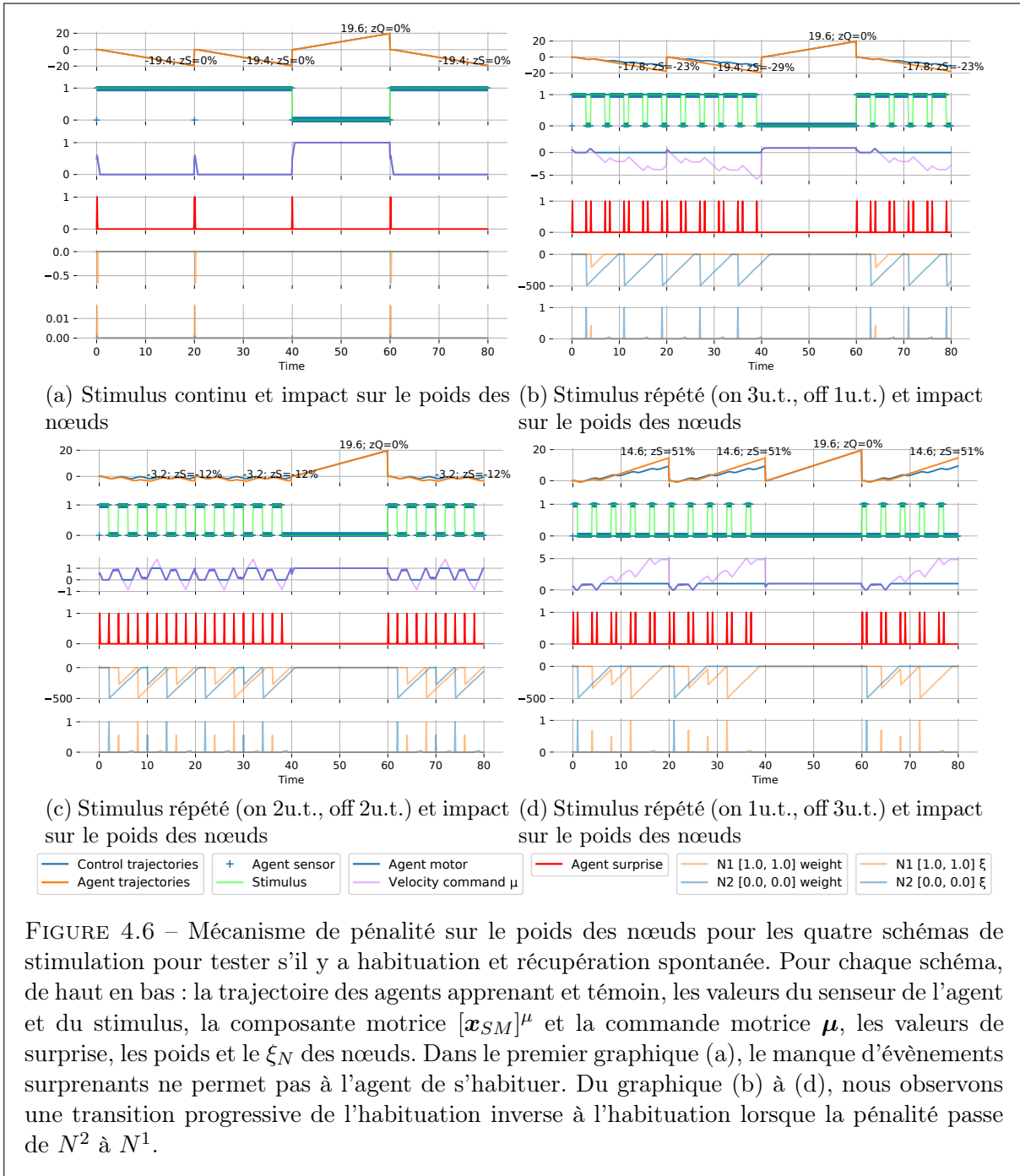
Chaque figure est composée de quatre sous-figures, une pour chaque schéma de stimulation. Chaque sous-figure a la même disposition. Le graphique du haut affiche la trajectoire de l’agent apprenant et la trajectoire de l’agent témoin. Sur la trajectoire de l’agent apprenant, il y a deux nombres : la dernière position de l’agent  $\mathbf{x}_E$  pour cette phase et un score relatif à l’agent témoin. Deux scores sont calculés à partir de la dernière position de l’agent témoin  $\mathbf{x}_c$  et de sa "meilleure" dernière position  $\mathbf{x}_c^*$  (la plus haute position atteinte par l’agent témoin, obtenue en l’absence de stimulation). Le score de la phase de stimulation, qui mesure l’efficacité de l’habituation et de la récupération spontanée, est défini par :

$$z_S(\mathbf{x}_E) = \frac{\mathbf{x}_E - \mathbf{x}_c}{\mathbf{x}_c^* - \mathbf{x}_c} \quad (4.14)$$

Ce score se situe dans l’intervalle  $[-\infty; 1]$ , avec  $z_S = 0$  si l’agent se comporte comme l’agent de contrôle et  $z_S = 1$  si l’agent n’est pas influencé par le stimulus, donc présente une habituation. Lorsque le score  $z_S$  est négatif, l’agent descend plus que l’agent témoin, il y a une sorte









d'habituation à la désactivation du stimulus, donc au comportement ascendant à la place du comportement descendant. Nous appelons ce phénomène *habituation inverse*.

Le score  $z_Q$  est le score pour la phase de calme, il mesure l'effet "rémanent" de l'habituation inverse (l'impact de l'habituation inverse se prolonge au début de la phase de calme, retardant le moment où l'agent commence à monter) et est défini par :

$$z_Q(\mathbf{x}_E) = \frac{\mathbf{x}_E - \mathbf{x}_c^*}{\mathbf{x}_c^*} \quad (4.15)$$

Ce score se situe dans l'intervalle  $[-1; 0]$ , avec  $z_Q = 0$  si l'agent termine exactement à la même position que celle de l'agent témoin et  $z_Q$  négatif s'il ne monte pas autant que l'agent témoin ( $\mathbf{x}_E < \mathbf{x}_c^*$ ). Dans ce dernier cas, cela indique que l'habituation inverse a un effet rémanent.

Le deuxième graphique en partant du haut donne des informations sur la perception du stimulus par l'agent en affichant la valeur du stimulus et la valeur du senseur de l'agent dans l'espace sensorimoteur (c'est-à-dire la seconde composante de  $\mathbf{x}_{SM}$ ). Le troisième graphique en partant du haut montre la valeur normalisée du moteur de l'agent dans l'espace sensorimoteur (la composante motrice  $[\mathbf{x}_{SM}]^\mu$ ), et la commande de vitesse  $\boldsymbol{\mu}$ . Les derniers graphiques donnent plus d'informations en fonction du modèle (mémoire sensorielle, impact sur l'énergie ou sur le poids des nœuds).

Pour l'agent témoin, le comportement de réponse (mouvement vers le bas) est observé avec l'activation du stimulus (phases 1, 2 et 4). Pendant la phase de calme, soit sans stimulation, on observe le comportement par défaut de l'agent témoin : il se déplace vers le haut.

L'habituation est observée avec chacun des trois modèles proposés, pour au moins un type de schéma de stimulation.

Pour le modèle avec mémoire sensorielle (section 4.1.1), comme le montre la figure 4.4, une activation plus longue du stimulus produit une plus forte habituation. Cela peut s'expliquer par le fait que la mémoire sensorielle  $\delta$  se rapproche de la moyenne de l'activation du stimulus (1, 0,75, 0,5 et 0,25 respectivement avec les schémas de stimulation **(a)**, **(b)**, **(c)** et **(d)**, voir les graphiques du bas des quatre sous-figures). Par conséquent, son effet d'inhibition est donc plus fort avec un temps d'activation plus long (voir l'équation (4.3) et la figure 4.2).

À l'inverse, les modèles avec pénalisation des nœuds (sections 4.1.1) présentent un effet d'habituation plus prononcé lorsque la durée d'activation du stimulus est strictement inférieure à sa durée d'inactivation (cas du schéma de stimulation **(d)**, voir les figures 4.5 et 4.6). Pour les deux modèles concernés, l'habituation dépend de l'occurrence des surprises provoquées par activation ou désactivation du stimulus. Pour le schéma de stimulation **(a)**, il y a trop peu de surprises pour que l'habituation se mettent en place. Pour les autres schémas de stimulation, le rapport entre la durée d'activation et d'inactivation a un impact sur l'habituation car il modifie l'intervalle de la dimension motrice couvert par la composante motrice  $[\mathbf{x}_{SM}]^\mu$  dans l'espace sensorimoteur.

Pour illustrer ce phénomène, nous considérons les trajectoires sensorimotrices d'un agent lors du schéma **(b)** avec un ratio d'activation du stimulus élevé (figure 4.7c) et lors du schéma **(d)** où celui-ci est faible (figure 4.7d). Dans le cas d'un ratio élevé d'activation du stimulus, l'agent passe

plus de temps "en haut" de l'espace  $SM$ , sous l'influence du nœud  $N^1$  qui amène la composante motrice  $[\mathbf{x}_{SM}]^u$  de l'agent proche de 0. Par conséquent, lorsque le stimulus est désactivé, l'agent va directement sur le nœud  $N^2$ , le pénalisant fortement. L'influence ainsi affaiblie de  $N^2$  combinée au temps d'inactivation plus court réduit la distance parcourue par l'agent "en bas" de l'espace  $SM$ , avant la réactivation du stimulus. Ainsi, même lorsque le stimulus est réactivé, l'agent n'est jamais assez proche de  $N^1$  pour que celui-ci soit pénalisé. Progressivement,  $N^2$  est de plus en plus pénalisé,  $N^1$  n'est pas affecté et asymptotiquement, l'agent oscille dans l'espace sensorimoteur entre des positions à composante motrice nulle (et donc un moteur dénormalisé à  $-1$ ). Ceci correspond à une situation d'habituation *inverse* où le comportement par défaut (vers le haut) est inhibé.

Dans la figure 4.7d avec un faible ratio d'activation du stimulus, la première activation et la première désactivation du stimulus sont les mêmes et ont les mêmes impacts sur les nœuds. Ensuite, bien que le nœud  $N^2$  subisse une forte pénalité lors de sa première "visite" par l'agent, le temps d'inactivation plus long est suffisant pour que l'agent passe à  $m = 1$ . Les mêmes arguments que précédemment avec des conditions opposées (temps d'activation plus court et forte pénalisation de  $N^1$ ) expliquent pourquoi l'agent oscille asymptotiquement dans l'espace sensorimoteur entre des positions dont la composante motrice est égale à 1. Ceci correspond à l'habituation où le comportement de réponse (vers le bas) est inhibé.

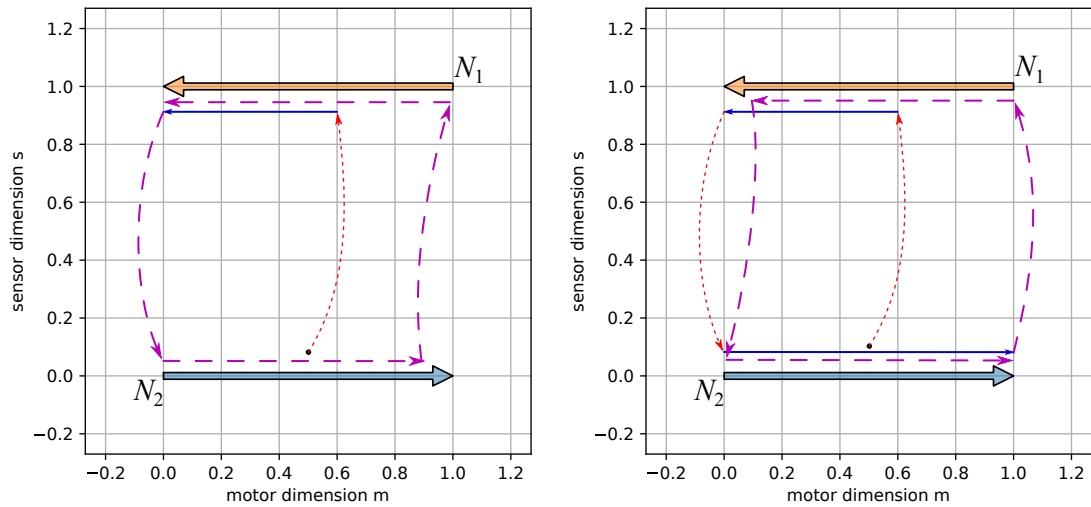
Considérons maintenant la caractéristique de récupération spontanée. Pour les trois modèles, l'évaluation de la récupération spontanée implique les deux dernières phases du protocole (la phase de calme suivie de la dernière phase de stimulation). La récupération spontanée est censée se produire pendant la phase de calme et son impact est évalué pendant la phase de stimulation suivante (phase 4). Dans tous les cas, les trajectoires des agents durant leur quatrième phase sont similaires à celles de leur première phase. Ceci s'explique par le fait qu'à la fin de la troisième phase, les composantes fondamentales du modèle considéré (mémoire sensorielle, énergie des nœuds ou poids des nœuds) reviennent à des valeurs proches de celles du début de la première phase.

En conclusion de cette section, nous constatons que les trois modèles proposés présentent une habituation et une récupération spontanée. Le rapport de stimulation a un impact sur l'amplitude de l'habituation. Cet impact n'est pas le même pour les trois modèles. En particulier, l'activation continue ne déclenche l'habituation que dans le modèle avec mémoire sensorielle. Enfin, il est intéressant de noter que le modèle de mémoire sensorielle, par l'intermédiaire de  $\beta$ , peut ajuster la vitesse de l'habituation et de la récupération spontanée, mais pas séparément. Plus  $\beta$  est élevé, plus l'habituation et la récupération spontanée sont rapides. Sur les modèles avec pénalité sur les nœuds, les deux caractéristiques peuvent être régulées indépendamment en ajustant leurs paramètres  $\gamma$  (pour réguler l'habituation) et  $\beta$  (pour réguler la récupération spontanée).

### Configuration pour l'intervalle interstimulus et résultats

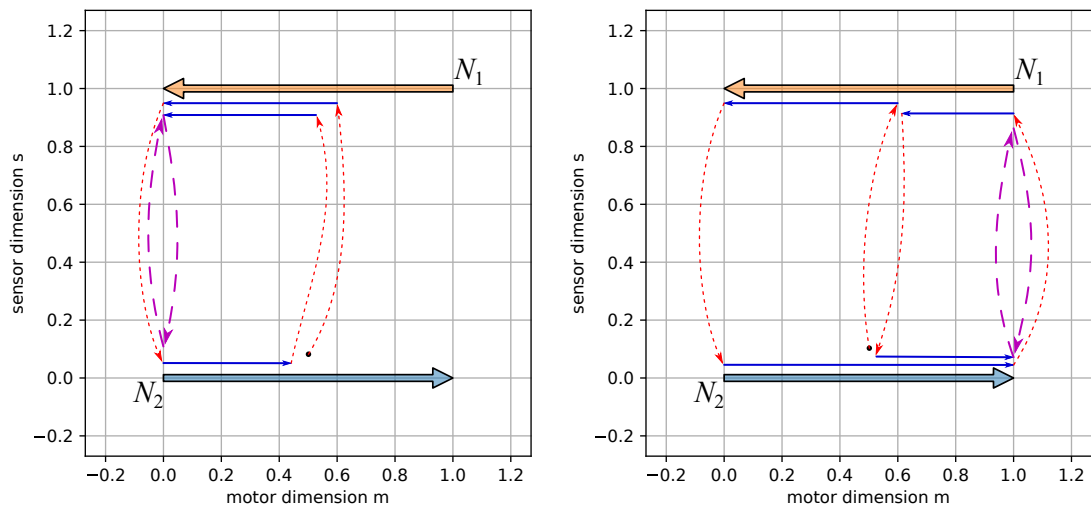
Dans cette section, nous nous concentrons sur l'effet de l'intervalle interstimulus qui est la quatrième caractéristique de l'habituation telle que décrite dans (Catharine H RANKIN et al., 2009). Selon les biologistes, il existe trois conséquences potentielles d'une stimulation à plus haute fréquence : un décrétement plus rapide de la réponse au stimulus, un décrétement plus prononcé de cette réponse et une récupération spontanée plus rapide.

Nous testons cette caractéristique en utilisant un protocole en deux phases. Pendant la pre-



(a) Agent témoin lors du schéma (b)

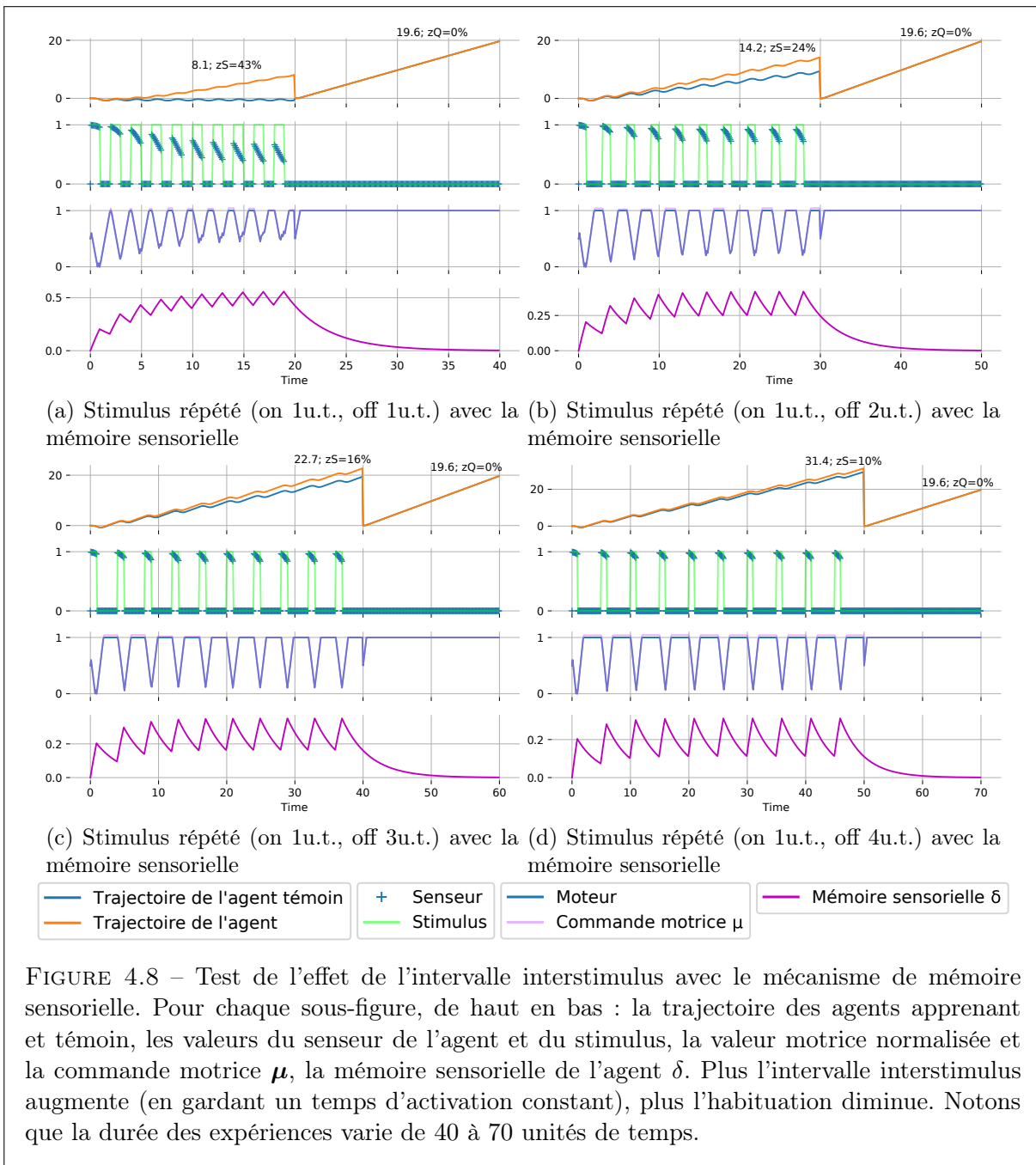
(b) Agent témoin lors du schéma (d)



(c) Agent avec pénalité sur les nœuds lors du schéma (b)

(d) Agent avec pénalité sur les nœuds lors du schéma (d)

FIGURE 4.7 – Trajectoires sensorimotrices montrant l’impact du rapport entre l’activation et l’inactivation du stimulus. Les flèches épaisses du haut et du bas représentent les deux nœuds. Les flèches horizontales fines représentent la trajectoire de l’agent dans l’espace sensorimoteur, elles devraient toujours être soit à  $s = 1$  lorsque le stimulus est activé, soit à  $s = 0$  lorsque le stimulus est désactivé, mais sont espacées pour une meilleure lisibilité. Les flèches verticales courbes montrent le saut dans la trajectoire dû à l’activation ou à l’inactivation du stimulus. Les flèches avec de grands tirets représentent le cycle dans lequel l’agent termine. L’agent commence toujours au point noir (également espacé verticalement pour plus de lisibilité). Les graphiques du haut représentent les trajectoires de l’agent témoin, ceux du bas sont les trajectoires de l’agent avec un modèle de pénalité des nœuds. Le graphique en bas à droite illustre l’habituation, tandis que le graphique en bas à gauche illustre l’habituation *inverse*.



mière phase, nous activons le stimulus dix fois, chaque activation étant suivie d'un intervalle interstimulus. Nous avons ensuite une phase de calme pour évaluer la récupération spontanée. La valeur de l'intervalle interstimulus est fixée à 2, 3, 4 ou 5 unités de temps. La durée du stimulus est toujours de une unité de temps.

Avec le modèle de mémoire sensorielle (voir figure 4.8), le comportement de l'agent montre une diminution de la réponse plus prononcée lorsque la fréquence de stimulation est plus élevée (score  $z_S = 43\%$  à la fréquence la plus élevée,  $z_S = 10\%$  à la fréquence la plus basse). Avec ce protocole, il n'est pas facile de dire si cette diminution de réponse est plus rapide ou non

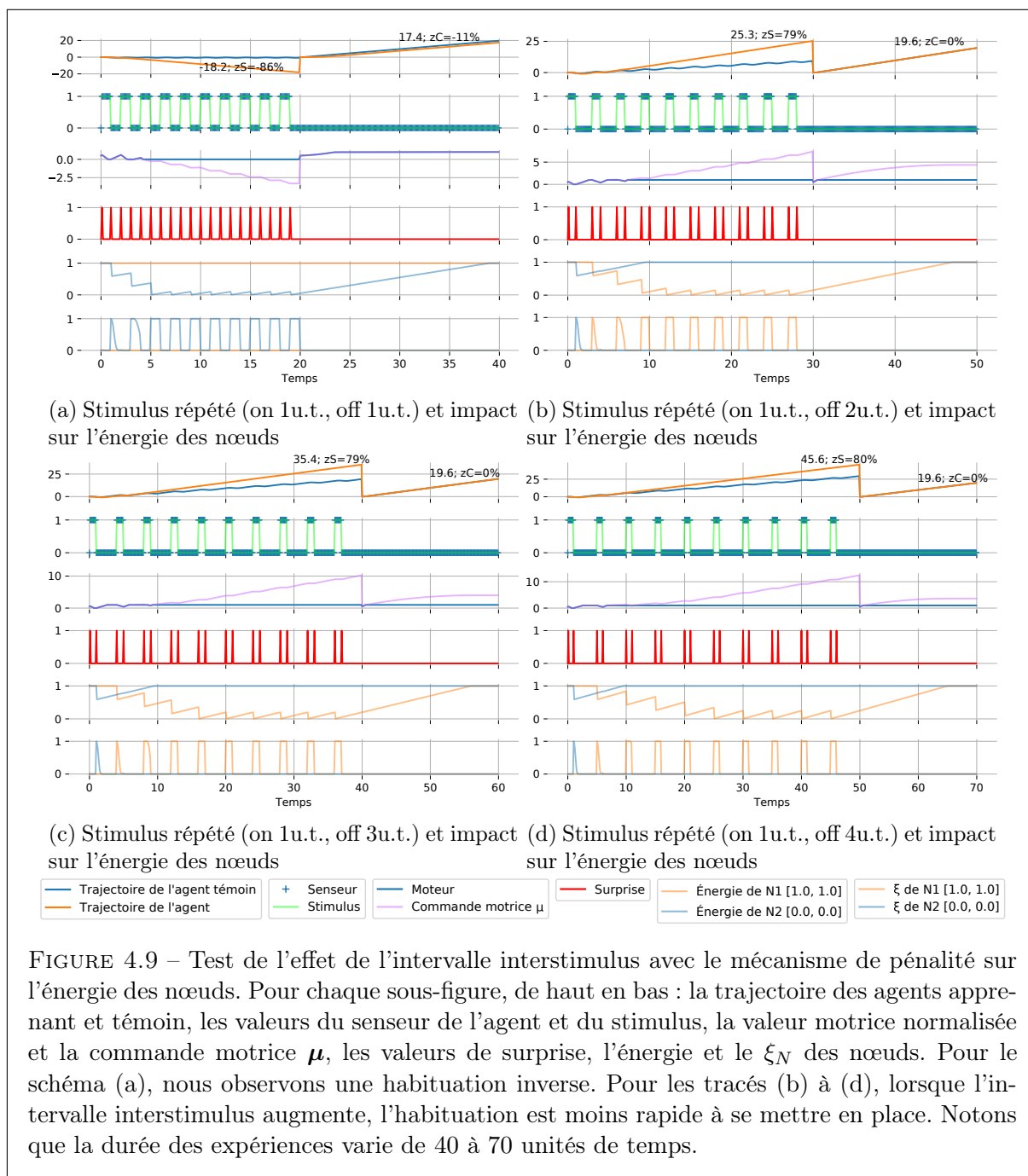
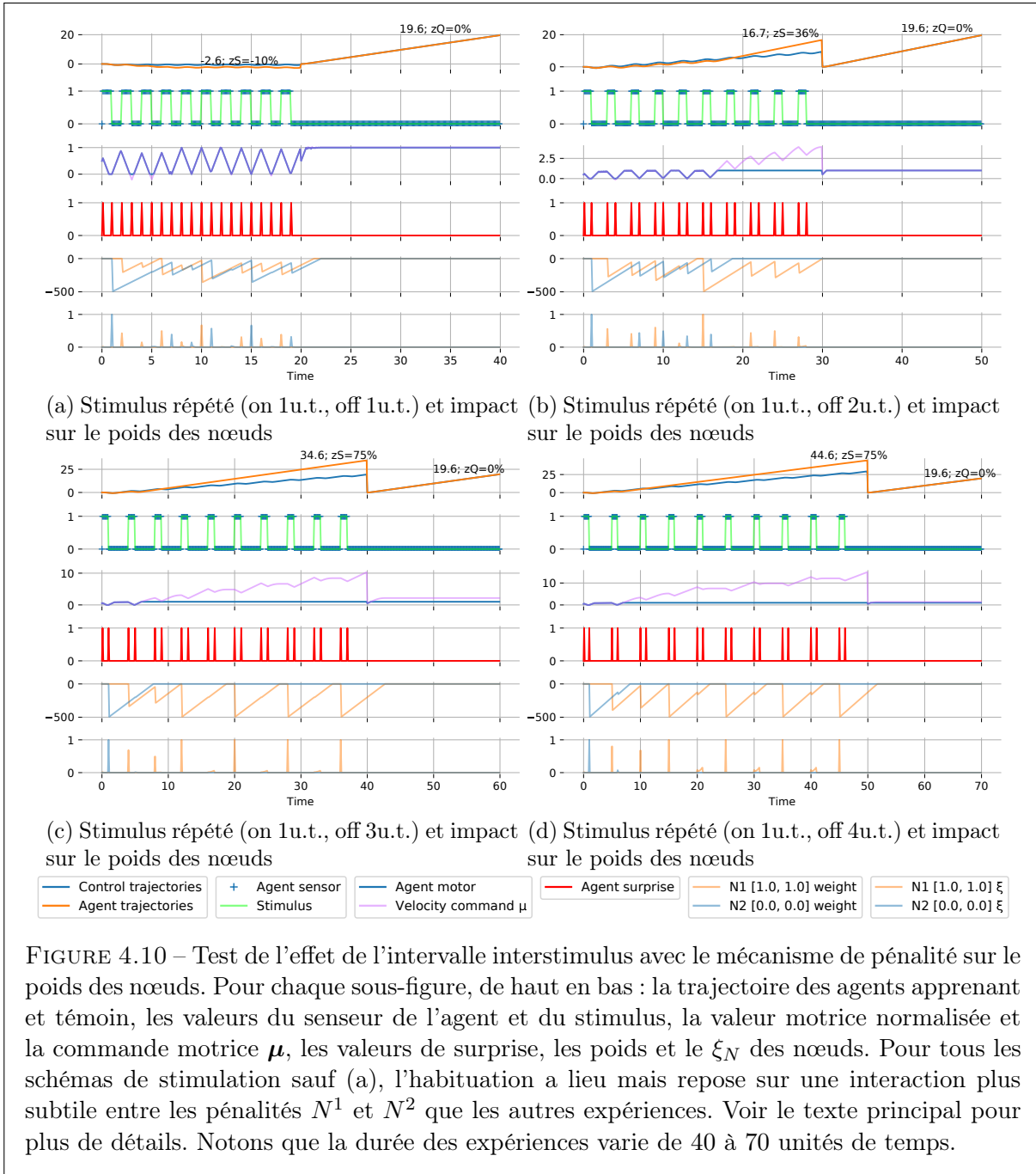


FIGURE 4.9 – Test de l'effet de l'intervalle interstimulus avec le mécanisme de pénalité sur l'énergie des nœuds. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du capteur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , les valeurs de surprise, l'énergie et le  $\xi_N$  des nœuds. Pour le schéma (a), nous observons une habituation inverse. Pour les tracés (b) à (d), lorsque l'intervalle interstimulus augmente, l'habituation est moins rapide à se mettre en place. Notons que la durée des expériences varie de 40 à 70 unités de temps.



du point de vue du comportement. Comme nous avons accès aux valeurs internes, nos analyses peuvent aller au-delà de ce niveau comportemental. Si l'on considère la mémoire sensorielle  $\delta$ , de la fréquence la plus élevée à la plus basse, il faut respectivement 15.9, 14.8, 14.1 et 13.6 unités de temps après la dernière désactivation du stimulus pour que  $\delta$  passe respectivement de 0.56, 0.42, 0.35 et 0.31 à moins de 0.01. Avec des fréquences plus élevées, la récupération spontanée demande plus de temps, mais son taux est plus rapide : respectivement 0,035, 0,028, 0,024 et 0,022. Ce taux de récupération  $r$  est calculé selon la formule suivante :

$$r = \frac{\delta_1 - \delta_2}{t} \quad (4.16)$$

avec  $\delta_1$  la valeur de  $\delta$  après la dernière désactivation du stimulus,  $\delta_2$  la première valeur inférieure à 0,01 et  $t$  le temps nécessaire à la diminution entre  $\delta_1$  et  $\delta_2$ . La tendance générale de  $\delta$  est d'augmenter jusqu'à osciller autour de la moyenne des valeurs brutes du senseur  $s_b$ . Cette valeur moyenne est plus élevée pour des fréquences de stimulation plus élevées. Les valeurs de l'intervalle interstimulus sont toujours trop courtes pour que  $\delta$  revienne à 0, elle s'accumule donc lors de chaque activation de stimulus. La quantité accumulée pendant une unité de temps d'activation dépend de la différence entre la valeur de départ et la valeur moyenne asymptotique. Une manière grossière d'évaluer l'effet d'habituation plus rapide consiste à mesurer le nombre de cycles d'activation par unité de temps, ou encore le temps nécessaire pour que  $\delta$  atteigne sa valeur moyenne asymptotique. Dans les deux cas, des fréquences plus élevées conduisent à une habituation moins rapide avec ce modèle : respectivement, des fréquences les plus élevées aux fréquences les plus faibles, 8, 8, 6, 7, 4, 7 et 0, 9 unités de temps sont nécessaires pour atteindre les valeurs moyennes respectives (0, 5, 0, 33, 0, 25 et 0, 2), soit le cinquième, le troisième, le deuxième et le premier cycle d'activation.

Pour le modèle de pénalité sur l'énergie des nœuds (voir figure 4.9), à l'exception du cas unique d'habituation *inverse* dans le schéma de stimulation **(a)**, il y a une diminution de réponse plus rapide, mais pas significativement plus prononcée puisque le score  $z_S$  est à peu près égal à 80%. En effet, avec les intervalles interstimulus des figures 4.9b, 4.9c et 4.9d, le niveau asymptotique de l'énergie du nœud  $N^1$  est atteint à la quatrième activation du stimulus. La diminution de la réponse est donc plus rapide car cette quatrième activation survient plus tôt pour des intervalles interstimulus plus courts. Dans la phase de calme, la récupération spontanée n'est jamais plus rapide, elle est exactement la même (soit 20 unités de temps) pour tous les intervalles interstimulus. Cela est attendu car le taux de récupération d'énergie est constant.

Dans le cas du modèle avec pénalité sur le poids des nœuds (voir figure 4.10), avec une fréquence décroissante, l'agent passe d'une habituation *inverse* ( $z_S = -10\%$ ) à une habituation de plus en plus prononcée ( $z_S = 75\%$  à la fréquence la plus basse). Lors de la phase de calme, la récupération spontanée n'est jamais plus rapide, car comme pour le taux de récupération d'énergie, le taux de récupération du poids est constant. La conséquence de la fréquence sur la diminution plus rapide de la réponse est plus complexe à analyser. Plus la fréquence est basse, plus la trajectoire sensorimotrice de l'agent est confinée au côté droit (composante motrice proche de 1) de l'espace sensorimoteur. Le nœud  $N^1$  est inhibé plus tôt dans le cas de la haute fréquence, mais la pénalité est plus forte dans le cas de la basse fréquence. L'effet global sur la vitesse de diminution de la réponse est difficile à évaluer.

Globalement, nos modèles ne montrent jamais l'effet complet de l'intervalle interstimulus. La récupération spontanée plus rapide n'est jamais observée, sauf si l'on considère son rythme plus rapide avec le modèle utilisant une mémoire sensorielle. En fait, dans nos modèles, l'effet de

l'intervalle interstimulus dépend surtout de la récupération spontanée, et en particulier de la façon dont elle a lieu : de façon permanente (cas des modèles de pénalisation des nœuds), ou directement après le dernier stimulus perçu (cas du modèle avec mémoire sensorielle).

## Combinaison des modèles

Chacun de nos modèles a ses propres propriétés, qui permettent à l'agent de s'habituer plus ou moins efficacement en fonction du schéma de stimulation (voir section 4.1.2). Mais ces modèles simples ne résolvent pas tous les cas : avec des paramètres fixes, il y aura toujours une situation où l'habituation sera plus difficile à réaliser. Cette section explore certaines combinaisons de nos modèles pour obtenir une habituation observée pour tous les schémas de stimulation. Ces combinaisons sont d'autant plus faciles à réaliser que nos mécanismes d'habituation sont complémentaires puisqu'ils ont un impact sur des composants indépendants les uns des autres. Cependant, même si l'on peut facilement construire des combinaisons de modèles, il n'y a pas de garantie que les propriétés soient conservées : par exemple, dans une combinaison avec un des modèles utilisant le mécanisme de surprise et le modèle utilisant la mémoire sensorielle, la surprise dépend de la valeur perçue du stimulus qui est affectée par la mémoire sensorielle.

Pour ces expériences, nous combinons le mécanisme de mémoire sensorielle avec chacun des modèles de pénalisation des nœuds (voir figures 4.11 et 4.12). Le premier résultat observé est qu'il n'y a pas de perte : l'habituation résultant d'une telle combinaison est toujours similaire ou plus prononcée que l'habituation des mêmes modèles utilisés individuellement. En effet, le score d'habituation  $z_S$  des figures 4.11 et 4.12 est toujours supérieur ou égal aux scores d'habituation des figures 4.4, 4.5 et 4.6.

En outre, dans les cas où un mécanisme de pénalisation des nœuds seul entraînerait une habituation *inverse* ( $z_S$  négatif, avec les schémas de stimulation **(b)** ou **(c)**), le modèle combiné entraîne une habituation "normale". Cela s'explique principalement par une asymétrie de la valeur de la surprise, qui est plus importante lors de l'activation du stimulus que lors de sa désactivation. Avec notre définition de la surprise (voir équation 4.5) et notre implémentation de la prédiction de l'état suivant (voir équation 4.6), et comme illustré sur la figure 4.13, la surprise est égale à la différence entre deux valeurs successives du senseur (représentée par une croix sur la figure). En s'accumulant pendant l'activation du stimulus, la mémoire sensorielle  $\delta$  entraîne une diminution de la valeur du senseur habitué  $s_h$  (représentée par une croix). Par conséquent, la valeur de surprise est plus faible à la désactivation du stimulus qu'à l'activation du stimulus. Le nœud  $N^2$  est moins pénalisé que le nœud  $N^1$  car le nœud  $N^2$  est impacté par une surprise plus faible lors de la désactivation alors que le nœud  $N^1$  est impacté par la surprise plus forte lors de l'activation. Par conséquent, le mécanisme de pénalisation des nœuds favorise l'habituation au lieu de l'habituation *inverse*.

### 4.1.3 Discussion autour de nos modèles d'habituation

Dans cette section, nous discutons d'abord de certaines différences entre nos modèles. Nous donnons ensuite notre avis sur les caractéristiques de l'habituation qui n'ont pas été abordées dans la section des résultats et dont l'étude pourrait être intéressante dans deux directions : soit pour les modélisateurs qui souhaiteraient étendre notre travail, soit parce que ces propriétés pourraient également être fondamentales pour les agents artificiels avec comportements adaptatifs.



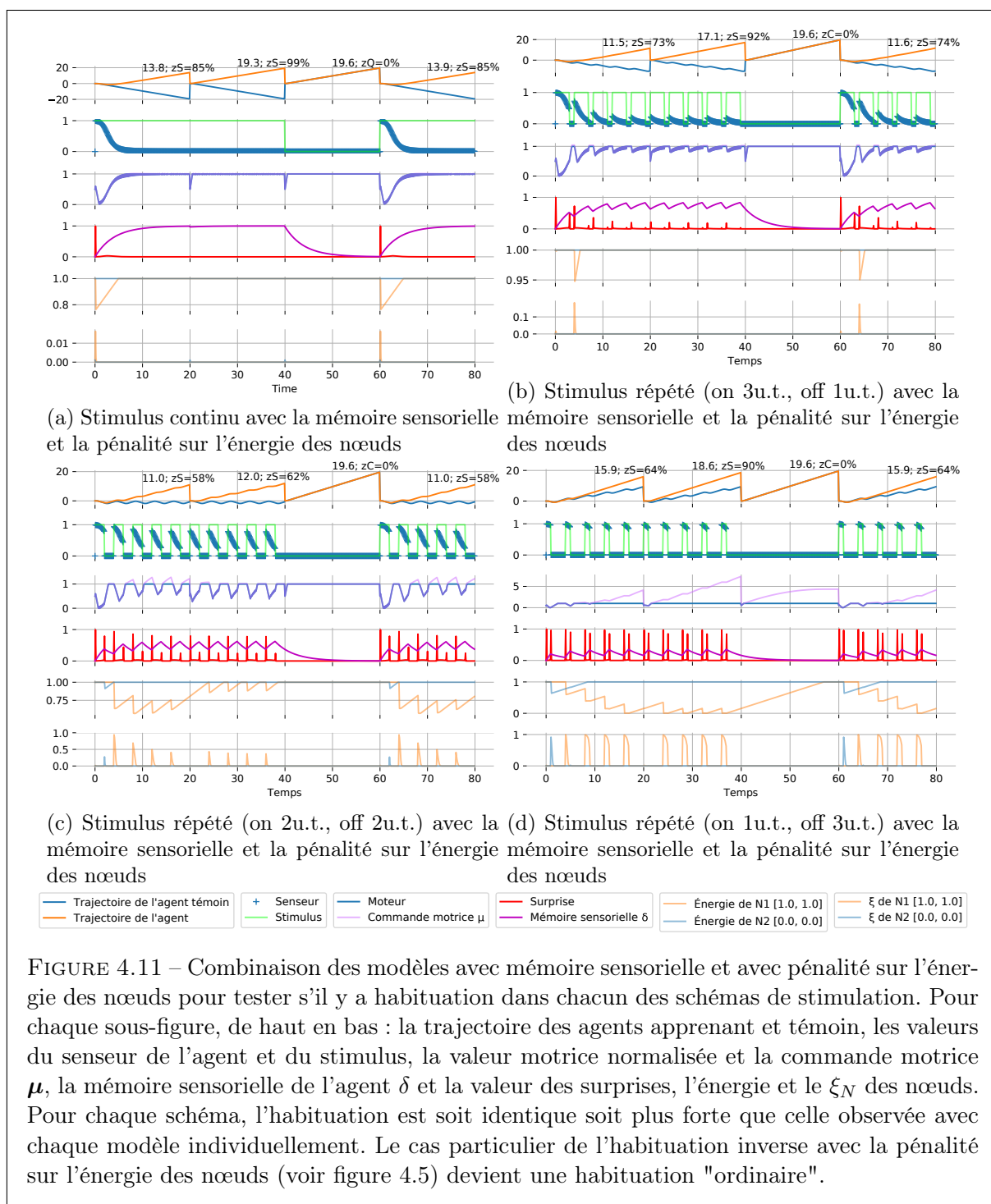


FIGURE 4.11 – Combinaison des modèles avec mémoire sensorielle et avec pénalité sur l'énergie des nœuds pour tester s'il y a habituation dans chacun des schémas de stimulation. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. Pour chaque schéma, l'habituation est soit identique soit plus forte que celle observée avec chaque modèle individuellement. Le cas particulier de l'habituation inverse avec la pénalité sur l'énergie des nœuds (voir figure 4.5) devient une habituation "ordinaire".

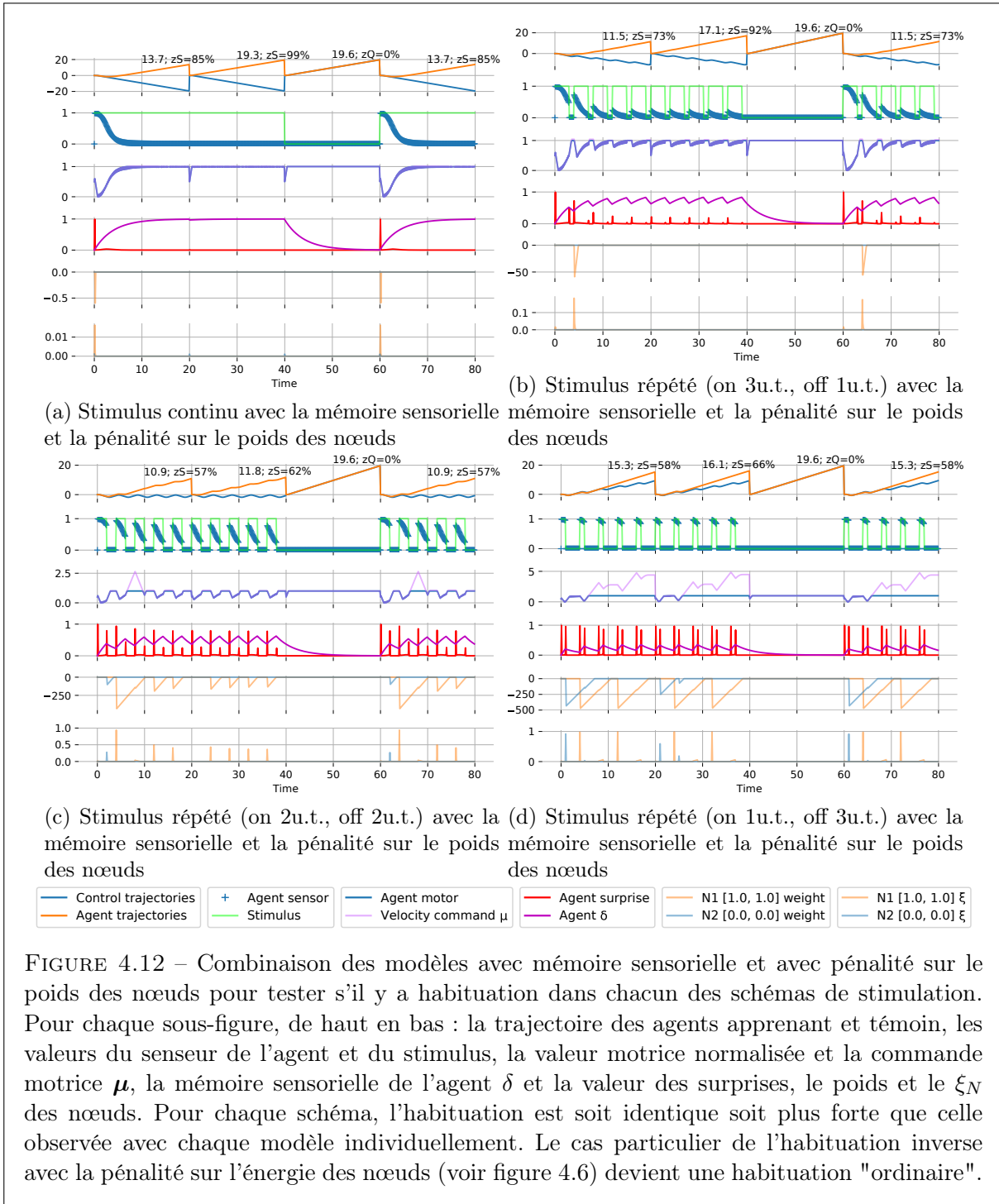
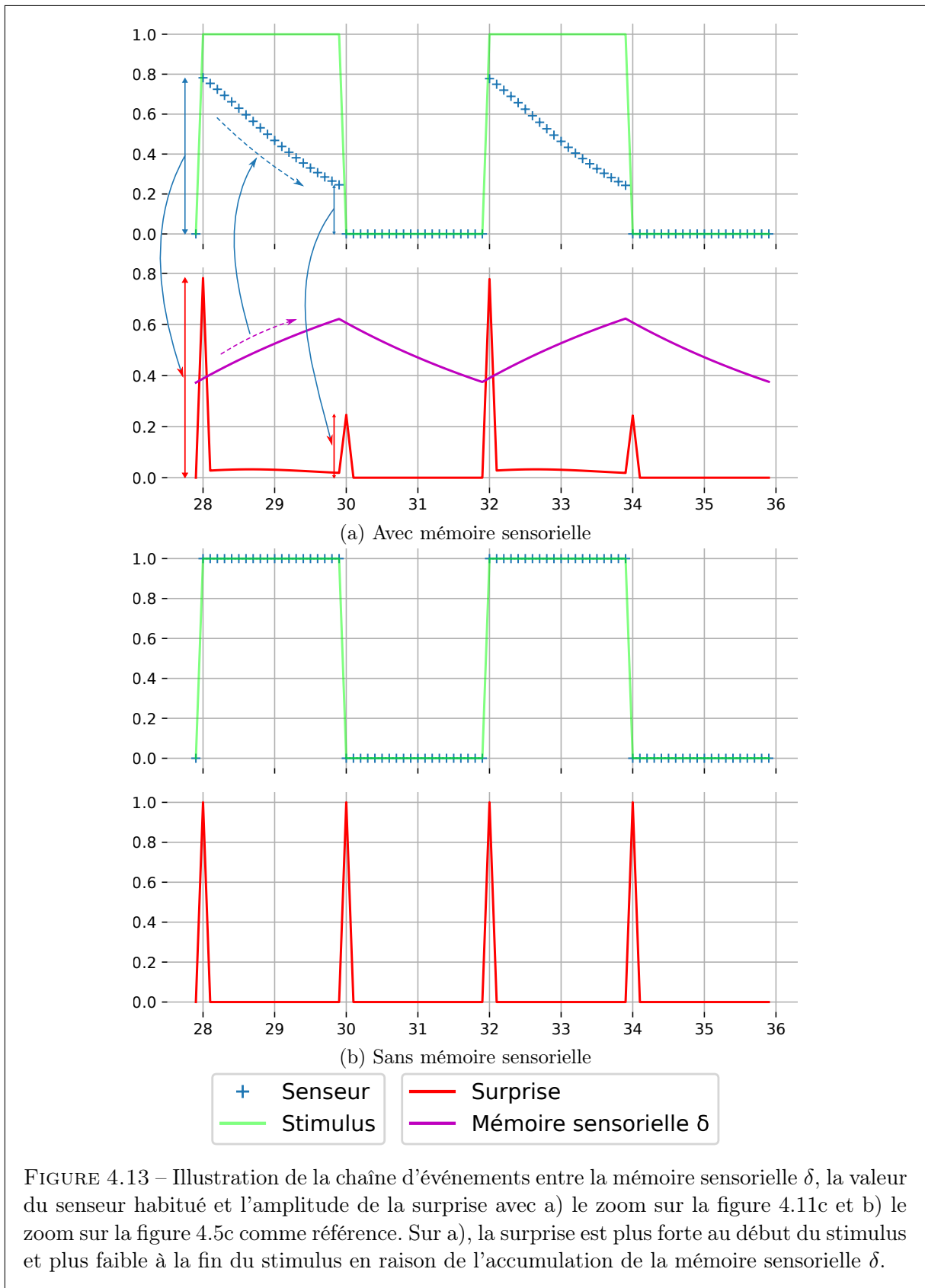


FIGURE 4.12 – Combinaison des modèles avec mémoire sensorielle et avec pénalité sur le poids des nœuds pour tester s’il y a habituation dans chacun des schémas de stimulation. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l’agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l’agent  $\delta$  et la valeur des surprises, le poids et le  $\xi_N$  des nœuds. Pour chaque schéma, l’habituation est soit identique soit plus forte que celle observée avec chaque modèle individuellement. Le cas particulier de l’habituation inverse avec la pénalité sur l’énergie des nœuds (voir figure 4.6) devient une habituation "ordinaire".



## Comparaison des modèles

Nos modèles sont conçus au niveau abstrait de l'IDSM. En cela, ils sont différents de la plupart des travaux connexes cités dans l'état de l'art (MARSLAND, 1999, WANG et HSU, 1990, BI et al., 2017, HONG et al., 2020) qui travaillent au niveau des neurones ou des synapses. Un léger avantage est que nos modèles, étant plus conceptuels, peuvent être plus faciles à analyser et à comprendre. Cependant, comme nous le verrons plus loin, une compréhension approfondie de nos modèles peut encore être complexe. D'autre part, beaucoup de ces travaux connexes sont plus avancés que les nôtres en termes d'application pratique : contrairement à (Stephen MARSLAND, NEHMZOW, SHAPIRO et al., 2000), nous sommes encore loin d'utiliser notre IDSM avec l'habituation dans un cadre robotique.

**Différences entre mémoire sensorielle et pénalité sur les nœuds** Les différents modèles que nous proposons ont été conçus pour permettre l'habituation en réponse à des stimulations différentes : la mémoire sensorielle spécifiquement face à un stimulus continu et la pénalité sur les nœuds spécifiquement face à un stimulus répété. Mais certaines différences vont au-delà.

Tout d'abord, alors que les modèles avec pénalité sur les nœuds ont un impact sur la dynamique sensorimotrice, le modèle de mémoire sensorielle modifie l'état sensorimoteur sans modification des nœuds (donc sans modification de la dynamique sensorimotrice). Avec la mémoire sensorielle, nous changeons artificiellement la position sensorielle de l'agent dans l'espace sensorimoteur, ce qui implique ensuite un changement dans la réponse motrice. Dans un modèle avec pénalité sur les nœuds, l'influence des nœuds est d'abord affectée, puis cela affecte la dynamique, qui à son tour affecte la réponse motrice.

Une autre différence est qu'avec la mémoire sensorielle, l'habituation se produit en fonction de la **valeur de l'intensité** du stimulus, alors que pour les modèles avec pénalité sur les nœuds, à travers le mécanisme de surprise, l'habituation se produit en fonction du **changement de l'intensité** du stimulus. Par conséquent, les modèles avec pénalité sur les nœuds peuvent rendre compte de l'habituation en cas de diminution soudaine de l'intensité du stimulus, comme cela a été observé chez les larves de poisson zèbre qui montrent une habituation aux diminutions brusques de l'éclairage (voir (RANDLETT et al., 2019)), alors que le modèle avec mémoire sensorielle ne le pourrait pas. C'est également cette capacité des modèles avec pénalité sur les nœuds à être insensibles à la valeur absolue de l'intensité du stimulus qui leur permet de produire ce que nous avons appelé l'habituation *inverse*. Une autre conséquence pour les modèles avec pénalité sur les nœuds est que la récupération spontanée peut également se produire lorsque le stimulus est encore actif s'il est perçu constamment avec la même intensité et pas uniquement lorsque le stimulus est inactif. Ces deux caractéristiques peuvent être considérées comme un inconvénient en empêchant l'habituation de se produire, mais le fait d'avoir à notre disposition plusieurs mécanismes différents est un avantage pour nous : notre objectif n'est pas la modélisation biologiquement plausible de l'habituation mais bien la conception de différents outils permettant à un agent artificiel de découvrir progressivement ses possibilités d'actions dans son environnement.

Dans les modèles avec pénalité sur les nœuds, la composante motrice de l'état sensorimoteur de l'agent a une très forte influence sur l'amplitude de l'effet de l'habituation (la réponse de l'agent au stimulus provenant de nœuds avec une position spécifique sur la dimension motrice) alors qu'elle n'en a absolument aucune pour le modèle avec mémoire sensorielle. Il n'est pas évident pour nous de savoir si, en biologie, l'apprentissage de l'habituation repose uniquement sur l'entrée sensorielle ou s'il dépend également de ce que fait l'organisme. Nous envisageons

la possibilité d'utiliser uniquement les composantes sensorielles dans les calculs de pénalité des nœuds dans des travaux futurs.

Enfin, comme nous l'avons dit à la fin de la section 4.1.2, le modèle avec mémoire sensorielle ne peut ajuster la vitesse d'habituation et la vitesse de récupération spontanée que conjointement, avec un paramètre unique, alors que les modèles avec pénalité sur les nœuds peuvent le faire séparément, à travers deux paramètres.

**Différences entre pénalité sur l'énergie et pénalité sur le poids** Il existe également des différences notables entre les deux modèles avec pénalité sur les nœuds. Avec les valeurs utilisées des paramètres, la perte d'énergie est plus progressive que la perte de poids, mais néanmoins, la perte d'influence des nœuds via l'énergie est plus importante que la perte d'influence des nœuds via le poids. Comme expliqué ci-dessous, cela est principalement dû à la distance à laquelle un nœud peut avoir de l'influence, distance qui est réduite lorsque le poids est diminué, car la distance à laquelle un nœud peut être pénalisé par une surprise est sa distance d'influence.

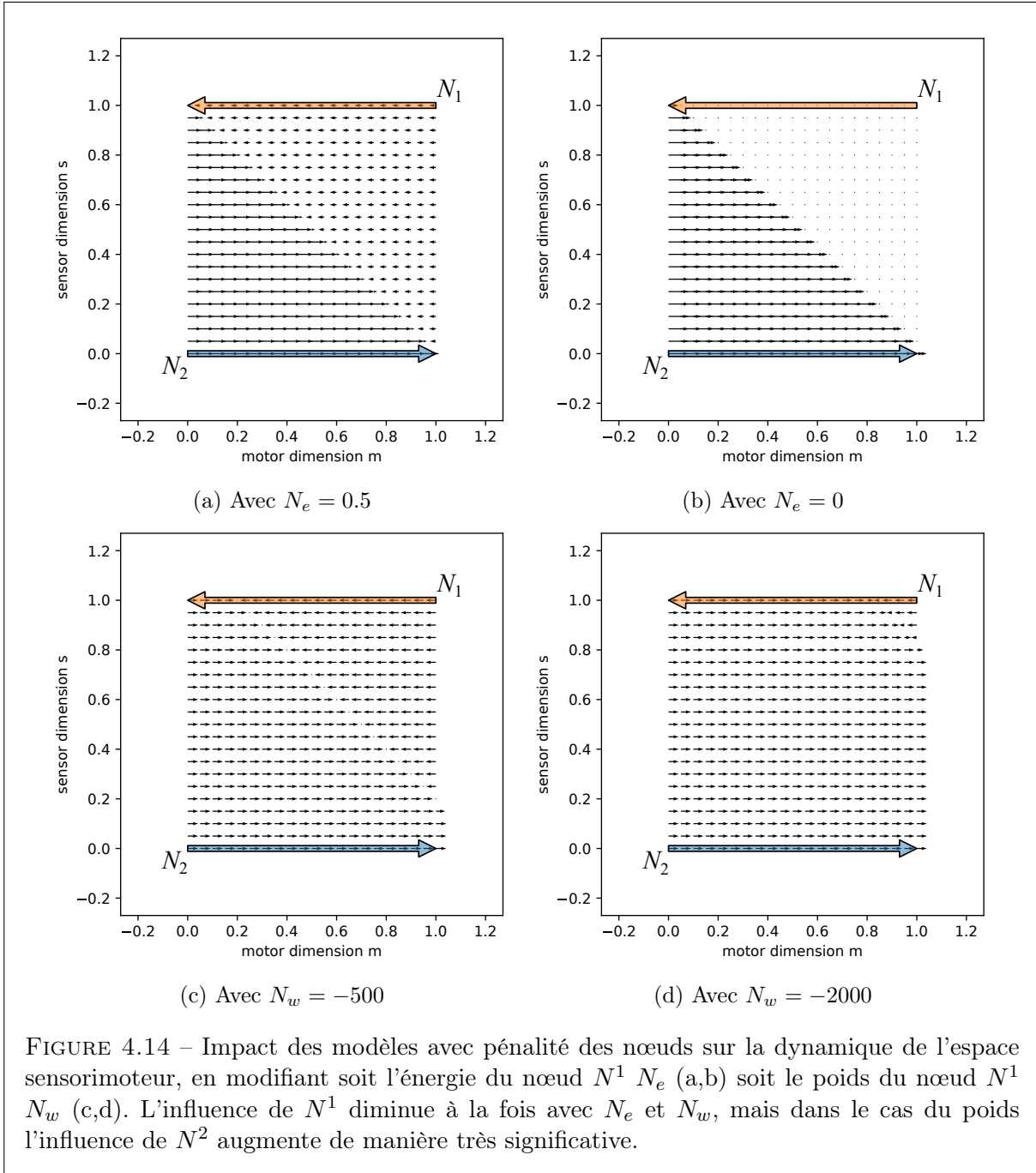
La principale différence est que dans le modèle pénalisant le poids, un poids plus faible signifie un  $\omega(N_w)$  plus faible (voir équation (3.3)). L'une des conséquences est qu'une diminution de  $\omega(N_w)$  entraîne une diminution de l'impact d'une même surprise à une distance donnée via  $\xi_N(\mathbf{x}_{SM})$  (voir les équations 4.12 et 4.13). Par conséquent, plus le poids du nœud est faible, plus il est difficile de l'inhiber davantage, puisque pour ce faire la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$  doit être de plus en plus proche de la position du nœud  $\mathbf{N}_p$ . Dans le modèle pénalisant l'énergie, un tel phénomène n'existe pas : la consommation  $c_e$  et la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$  étant égales par ailleurs, un nœud  $N$  subira la même pénalité énergétique indépendamment que  $N$  soit déjà affaibli ou non. De plus, pour une surprise donnée, la pénalité sera toujours significative même à une distance plus grande dans le modèle impactant l'énergie que dans celui impactant le poids, et ce même à l'initialisation (voir par exemple les premières pénalités dans le cinquième graphique en partant du haut dans les figures 4.5a et 4.6a).

Une autre différence significative est que le modèle pénalisant l'énergie diminue l'amplitude de l'influence du nœud alors que le modèle pénalisant le poids diminue la *portée* de l'influence du nœud, ce qui augmente celle des autres nœuds. Lorsque l'énergie diminue, la commande du nœud s'affaiblit, mais la frontière entre les influences des nœuds reste presque sur la médiatrice entre les deux nœuds (voir figures 4.14a et 4.14b). Avec un poids décroissant, la commande du nœud reste importante, mais la frontière entre les influences des nœuds se déplace significativement en faveur du nœud le plus fort (voir les figures 4.14c et 4.14d).

Malgré les deux différences "internes" mentionnées ci-dessus, il y a peu de différence observable au niveau comportemental dans nos expériences, si ce n'est que l'habituation (inverse ou non) est souvent plus prononcée pour le modèle pénalisant l'énergie. Cependant, il se pourrait que dans des contextes expérimentaux différents (avec un plus grand nombre de nœuds et une dynamique sensorimotrice plus complexe), elles provoquent des effets plus significatifs comme des comportements différents plutôt que seulement une différence d'expressivité d'un même comportement.

### Autres caractéristiques de l'habituation

Dans cette section, nous discutons brièvement des caractéristiques de l'habituation que nos modèles ne traitent pas explicitement et qui n'ont pas été testées expérimentalement.



**Potentialisation** Par définition de la potentialisation, des blocs répétés d'entraînement à l'habituation et de récupération spontanée conduisent à une habituation progressivement plus profonde (les effets de l'habituation sont plus prononcés) et plus rapide : (Catharine H RANKIN et al., 2009).

Avec les modèles proposés dans ce manuscrit, la potentialisation ne peut être observée que lorsque la récupération n'est pas complète. Si la récupération est complète, par exemple avec une longue phase de calme, il n'y aura pas de potentialisation avec les modèles proposés car tous les composants de base seront revenus à leur valeur initiale. Si, par contre, la récupération n'est pas complète, par exemple avec une courte phase de calme, nous observerons une diminution de la réponse dès le début de la phase de stimulation suivante et par la suite plus prononcée que lors des phases de stimulation précédentes. Cela s'explique par le fait que les composants de base ne sont pas revenues à leurs valeurs initiales et qu'ils induisent encore une certaine habituation.

À noter que nous ne savons pas si cela correspond à la potentialisation telle que décrite par les biologistes. Avec les modèles proposés, la "vitesse" de l'habituation est constante, mais elle commence avec une réponse déjà habituée lorsque la période de récupération est courte. Les modèles proposés ne peuvent pas rendre compte de la potentialisation en tant qu'habituation à un rythme "plus rapide".

**Effet de l'intensité du stimulus** Une autre propriété de l'habituation observée par les biologistes concerne l'effet de l'intensité du stimulus, où un stimulus plus faible entraîne une diminution plus rapide ou plus forte de la réponse.

Il est difficile de savoir dans quelle mesure nos modèles pourraient rendre compte de cette observation. En fait, nous aurions besoin d'entrer dans plus de détails sur les installations utilisées par les biologistes pour présenter et quantifier cette observation.

**Habituation en dessous de zéro** L'habituation en dessous de zéro est définie par le fait que les effets de la stimulation peuvent continuer à s'accumuler même après que la réponse ait diminué jusqu'à devenir nulle ou asymptotique.

Tous nos modèles peuvent rendre compte de cette observation indirectement. Par définition (voir équation 3.11), la commande de vitesse  $\mu$  peut continuer à s'accumuler après l'inhibition complète de la réponse. Si elle n'est pas bridée dans  $[0, 1]$ , elle peut s'accumuler au-delà des limites de l'espace sensorimoteur  $SM$ , et si nous n'avons pas réinitialisé l'agent entre les phases, il faudra de plus en plus de temps pour que le comportement se rétablisse.

**Spécificité au stimulus** Dans notre configuration, l'agent ne dispose que d'un seul senseur. Nous le considérons comme un senseur spécifique au stimulus, capable de percevoir un stimulus donné d'intensité variable. Ainsi, afin d'évaluer la spécificité du stimulus de l'habituation, nous devrions instancier l'agent avec plusieurs senseurs. Avec le modèle de mémoire sensorielle, si l'agent a une mémoire pour chacun de ses senseurs, alors l'habituation montrerait une spécificité au stimulus. Pour les modèles avec pénalité sur les nœuds, cela dépendrait de la position des nœuds responsables du comportement de réponse au stimulus. Si de tels nœuds sont "partagés" entre les senseurs captant les stimuli (par exemple un seul nœud avec  $N_p = [x, 1, 1]$ ), toute pénalité provoquée avec l'un des stimuli impacte cet unique nœud et par conséquent la réponse à l'autre stimulus. Mais avec les paramètres tels qu'ils sont fixés, la proximité nécessaire

entre l'agent et les nœuds dans l'espace sensorimoteur pour provoquer des pénalités significatives implique des nœuds pour chaque dimension sensorielle et donc une spécificité au stimulus.

**Déshabituaiton** La déshabituaiton se manifeste par un renforcement de la réponse habituée lorsqu'un stimulus nouveau ou nocif est inséré dans une série de stimulation provoquant l'habituation. Pour que cela se produise, à l'inverse de la spécificité au stimulus, nous aurions besoin d'une certaine dépendance ou interaction entre les stimuli ou leurs senseurs respectifs. Ceci n'est pas mis en œuvre dans les modèles proposés.

**Habituation de la déshabituaiton** Comme il n'existe actuellement aucun mécanisme de déshabituaiton dans nos modèles, il n'y a pas non plus d'habituation de la déshabituaiton. Néanmoins, si cette caractéristique reposait sur l'inhibition de la perception du stimulus responsable de la déshabituaiton, la répétition dudit stimulus devrait effectivement déclencher l'habituation à ce stimulus et donc indirectement l'habituation de la déshabituaiton.

**Habituation à long terme** L'habituation à long terme, qui correspond à une diminution de la réponse conservée pendant des heures, des jours ou des semaines, peut être prise en compte en réglant les paramètres des modèles proposés. Pour les modèles avec pénalité sur les nœuds, l'habituation et la récupération peuvent être réglées indépendamment et par conséquent la récupération peut être étalée sur le temps voulu. Pour le modèle avec mémoire sensorielle, l'habituation et la récupération ont lieu sur la même échelle de temps, donc une récupération lente implique une habituation lente. Nous pouvons également combiner plusieurs modèles, et ainsi avoir plusieurs habituations et leurs récupérations associées à des échelle de temps différentes.

#### En utilisant l'IDSM déformable

$\Delta t$	$k_t$	$k_d$	$k_\omega$	$\tau$	$\mu$
0.1	1	30	0.025	10	$\in [0; 1]$

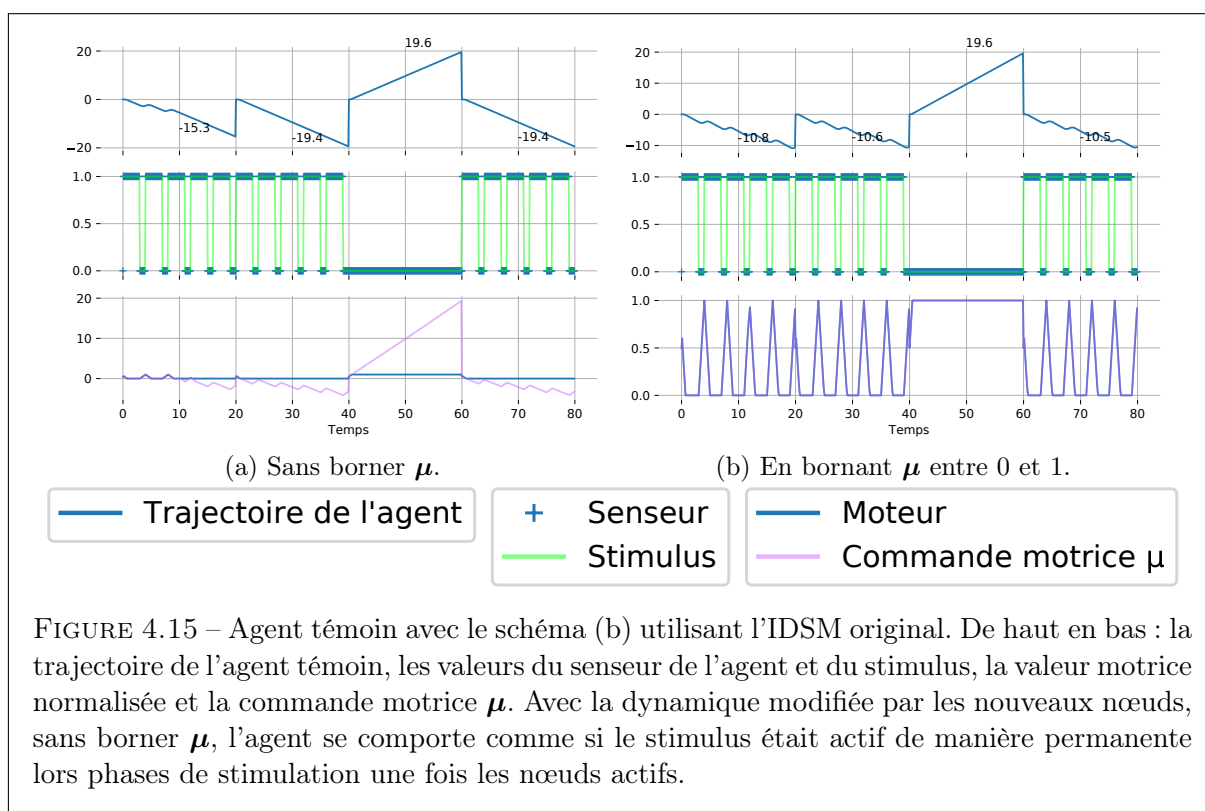
TABLE 4.2 – Valeurs des paramètres et options utilisées pour les résultats sur l'habituation avec IDSM déformable (section 4.1.3)

Dans nos modèles, nous faisons l'hypothèse forte que le poids des nœuds est maintenu constant en dehors des pénalités sur le poids et qu'aucun nouveau nœud n'est créé. Ces deux mécanismes ont été mis de côté dans notre étude afin de simplifier l'analyse de nos modèles. Notre travail peut donc être plutôt considéré comme étant basé sur une variante simplifiée de l'IDSM appelée "carte sensorimotrice basée sur des nœuds" ("*node based sensorimotor map*", NB-SMM, WOOLFORD et M. D. EGBERT, 2020).

Cependant, ces deux mécanismes sont critiques pour la partie déformable de l'IDSM qui permet l'encodage et le maintien des habitudes. Dans des travaux très préliminaires avec les modèles de mémoire sensorielle et de pénalité sur l'énergie des nœuds, nous observons toujours l'habituation avec à la fois la création de nœuds et la modification des poids comme dans l'IDSM original. La récupération spontanée est aussi observée malgré la création de nouveaux nœuds qui seraient susceptible de maintenir fortement le comportement "habitué" et de forcer l'agent dans une habitude d'ignorer le stimulus de manière permanente.

Lors de ces travaux, nous nous replaçons dans les mêmes conditions que lors de la section 4.1.2, en commençant avec les deux nœuds initiaux comme dans la figure 4.3. La seule différence,





hormis la création de nœuds et la mise à jour des poids des nœuds, est que nous contraignons la commande motrice  $\mu$  dans  $[0, 1]$ . Ce choix a été fait car l'activation des nouveaux nœuds modifie l'équilibre entre les nœuds initiaux et si  $\mu$  n'est pas bornée, il pourrait y avoir accumulation d'inertie et donc insensibilité de l'agent témoin aux activations et désactivations pendant les phases de stimulation. En l'état, il descend en permanence pendant ces phases avec le schéma (b) et il monte en permanence avec le schéma (d) (voir la figure 4.15 pour le schéma (b) donné en exemple). Par conséquent, mesurer l'habituation dans ces conditions donne des grands scores pour le premier schéma (b) et de très faibles scores pour le schéma (d). Contraindre  $\mu$  permet de retrouver les mêmes comportements de l'agent témoin que ceux de la section 4.1.2.

La figure 4.16 illustre l'application du modèle de mémoire sensorielle à un IDSM complet et nous observons toujours l'habituation et la récupération spontanée. Une analyse plus approfondie est nécessaire mais notre hypothèse sur la raison pour laquelle cela fonctionne est la suivante. En plus des deux premiers nœuds, de nouveaux nœuds sont créés sur la trajectoire de l'agent dans l'espace sensorimoteur. Le temps nécessaire pour qu'ils soient actifs ( $\tau = 10$  par défaut dans l'IDSM original) permet à la mémoire sensorielle de forcer la valeur sensorielle très proche de 0, donc à ce moment-là l'agent n'est proche que de nœuds renforçant le comportement ascendant (comportement par défaut hors stimulation). Après la phase de récupération, l'agent perçoit à nouveau pleinement le stimulus et revisite les nœuds renforçant le comportement de réponse. L'agent va alors vers le bas. Il est intéressant de voir que la trajectoire motrice est beaucoup plus rigide en début de dernière phase (les nouveaux nœuds sont actifs) qu'en début de première (les nouveaux nœuds ne sont pas encore actifs), le changement de comportement est beaucoup plus brusque, l'agent passe plus de temps avec la composante motrice  $[x_{SM}]^\mu$  à 0 ou à 1 et moins de temps dans la transition.

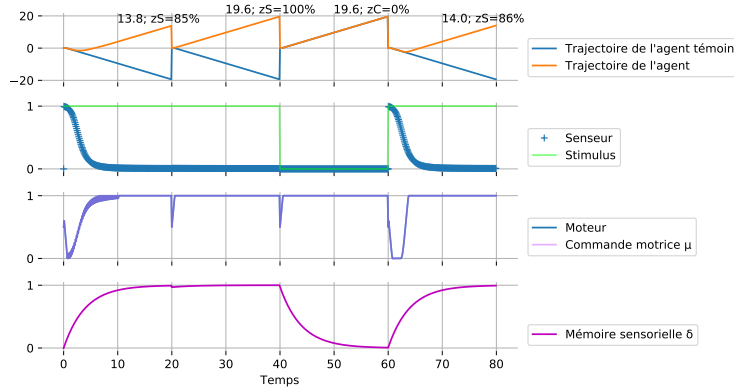


FIGURE 4.16 – Modèle avec la mémoire sensorielle basé sur l'IDSM original avec le schéma (a) pour tester s'il y a habituation. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du capteur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$ . L'habituation et la récupération spontanée sont toujours observées même si de nouveaux nœuds sont créés et les poids des nœuds continuellement mis à jour.

La figure 4.17 illustre le travail préliminaire sur l'utilisation du modèle avec pénalité sur l'énergie pour un IDSM complet avec le schéma (d). L'habituation et la récupération sont observables, et comme pour le modèle avec mémoire sensorielle une différence notable entre la trajectoire motrice de la phase 1 et celle de la phase 4 est attribuable au délai d'activation des nouveaux nœuds. En particulier, la trajectoire motrice de la phase 4 illustre bien l'habituation de la réponse, avec les cinq pics de réponse de plus en plus réduits.

Ces résultats préliminaires indiquent que le modèle avec pénalité sur l'énergie semble bien compatible avec l'IDSM complet mais il ne faut pas oublier que, dans nos expériences avec les modèles avec pénalités sur les nœuds, la composante sensorielle est binaire et reste soit 0 soit 1, sans valeurs intermédiaires. Dans une dernière expérience avec l'IDSM complet, nous mélangeons mémoire sensorielle et pénalité sur l'énergie, ce qui permet d'avoir la composante sensorielle variant continuellement dans  $[0, 1]$ .

Comme le montre la figure 4.18, cette combinaison présente de l'habituation et de la récupération spontanée, avec des scores en phases 1 et 2 légèrement meilleurs que lorsque le modèle avec énergie est utilisé seul. Le score de la phase 4 est un peu plus bas (43% à la place de 50%), probablement car il y a de plus nombreux nœuds répartis sur une plus grande zone sensorielle. La diminution des réponses est par contre toujours bien observée sur la trajectoire motrice.

Là encore, d'autres expériences et analyses approfondies devraient être menées pour évaluer la compatibilité des mécanismes de mémoire sensorielle et d'énergie avec l'IDSM "déformable", compatibilité qui ne doit pas être considérée comme acquise.

Le cas du modèle avec pénalité sur les poids est plus difficile à combiner avec un IDSM complet car il faudrait paramétrer soigneusement la modification continue des poids ainsi que les pénalités pour que l'habituation apparaisse sans conflits entre les deux mécanismes. Aucune de nos expériences préliminaires n'a été concluante.

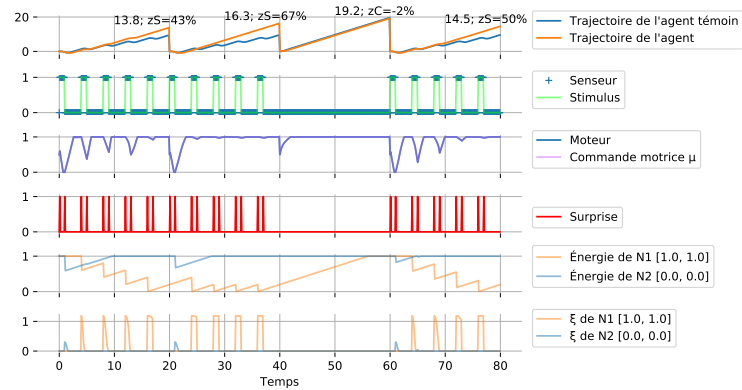


FIGURE 4.17 – Modèle avec pénalisation de l'énergie utilisant l'IDSM original avec le schéma (d) pour tester s'il y a habituation. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du capteur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. L'habituation et la récupération spontanée sont toujours observées avec la création de nouveaux nœuds et la modification continue des poids.

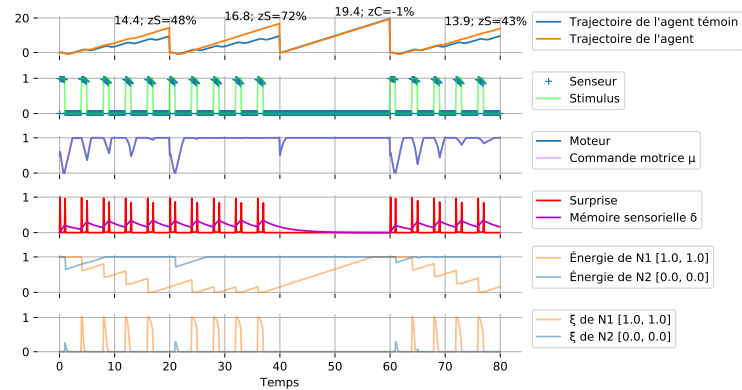


FIGURE 4.18 – Modèle avec mémoire sensorielle et pénalisation de l'énergie utilisant l'IDSM original avec le schéma (d) pour tester s'il y a habituation. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du capteur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. L'habituation et la récupération spontanée sont toujours observées avec la création de nouveaux nœuds, la modification continue des poids et la combinaison de deux modèles.

#### 4.1.4 Bilan sur l'habituatation

Dans cette section, nous avons proposé plusieurs extensions de l'IDSM pour rendre compte de certaines des multiples facettes de l'habituatation. Ces modèles sont concentrés sur les deux principales propriétés de l'habituatation : l'inhibition de la réponse et la récupération spontanée. Ils présentent des caractéristiques différentes et leur combinaison permet de couvrir davantage de cas que séparément. Les modèles proposés pourraient être considérés comme une base qui pourrait être affinée, combinée ou étendue pour rendre compte de toutes les propriétés de l'habituatation telles que définies par les biologistes.

Disposer d'un modèle opérationnel de l'habituatation, un mécanisme fondamental et répandu, permet également de s'interroger de manière pratique sur la façon dont l'habituatation contribue à l'adaptation d'un organisme à son environnement. Cette question n'est pas seulement intéressante d'un point de vue biologique mais aussi pour les agents artificiels. Dans le cadre de cette thèse, nous cherchons à permettre à un agent artificiel d'enrichir son répertoire comportemental à l'aide de mécanismes donnant de la plasticité. Le fait que nos modèles d'habituatation permettent à l'agent d'explorer son espace sensorimoteur de manière inhabituelle et de modifier la dynamique de celui-ci de manière différente qu'avec les mécanismes de renforcement de comportements déjà explorés de l'IDSM seul est une première réponse satisfaisante à cet enrichissement.

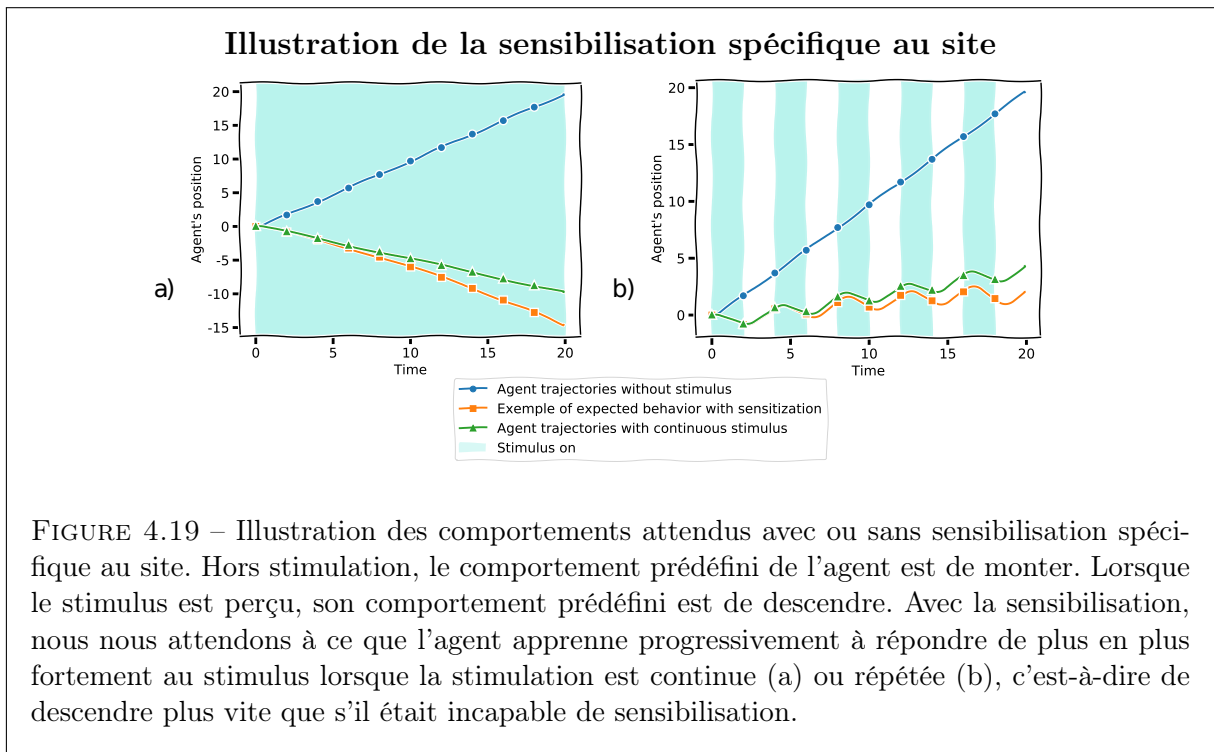
L'utilisation des mécanismes donnant à l'IDSM sa caractéristique "déformable", à savoir la création de nouveaux nœuds et la modification permanente des poids, est peu couverte ici mais une première approche est présentée et donne des résultats encourageants pour deux modèles. D'autres analyses doivent cependant être effectuées pour solidifier ces travaux. En particulier, nous pensons que l'apport de rigidité pourrait impacter la récupération spontanée, la rendant difficile si les nouveaux nœuds empêchent l'agent de sortir d'un comportement "habitué". Nous ne savons pas s'il est possible, selon la configuration, qu'il devienne insensible au stimulus de manière permanente. L'effet du paramètre  $\tau$  sur la période d'inactivation des nœuds après création nous semble aussi important lors du déploiement de la dynamique sensorimotrice en début d'expérience.

## 4.2 Sensibilisation spécifique au site

La sensibilisation spécifique au site (section 2.1.2) se définit, contrairement à l'habituatation, par une augmentation plutôt qu'une diminution de la réponse au stimulus lorsqu'il est répété. Nous adaptons nos trois modèles pour qu'ils puissent intégrer cet apprentissage. Ces travaux restent préliminaires et visent à traiter de la sensibilisation à part de l'habituatation. Des pistes seront données pour la combinaison d'apprentissages non-associatifs en fin de chapitre.

Pour les mêmes raisons que dans le cas de l'habituatation, le modèle avec mémoire sensorielle est plus adapté aux stimuli continus et les modèles avec bonus (plutôt que pénalité, dans le cadre de la sensibilisation) sur les nœuds sont plus adaptés aux stimuli brefs et répétés.

À nouveau, nous illustrons le comportement que nous attendons de nos modèles à l'aide d'une expérience de pensée, illustrée sur la figure 4.19. Cette fois, l'objectif est d'obtenir une augmentation de la réponse lorsqu'un stimulus est actif et un retour partiel à la réponse initiale lorsque le stimulus est absent. Nous utilisons la même configuration que pour l'habituatation : un agent simple, avec seulement un moteur et un senseur. Il existe un stimulus dans l'environnement qui peut être activé ou désactivé. Lorsque le stimulus est actif, l'agent le perçoit avec une valeur



constante quelle que soit la position de l'agent dans l'environnement. L'agent peut se déplacer le long d'une seule dimension, vers le haut ou vers le bas. Son comportement réflexe est de monter en l'absence de stimulus et de descendre lorsque l'agent perçoit le stimulus. Dans un tel contexte, une sensibilisation spécifique au site réussie devrait se traduire par le renforcement de la réponse (mouvement descendant) après une présentation continue ou répétée du stimulus : l'agent descend de plus en plus vite.

Dans la figure 4.19, nous considérons deux configurations et, pour chacune, nous montrons comment l'agent est censé se comporter en l'absence de stimulus (toujours la trajectoire du haut), en présence d'un stimulus en étant incapable de sensibilisation spécifique au site (toujours la trajectoire du milieu) et en présence d'un stimulus en étant capable de sensibilisation. Lorsqu'un stimulus est présenté, il peut être continu ou répété et, pour faciliter la lecture, les périodes où le stimulus est actif sont représentées par un fond coloré.

Dans la première configuration (figure 4.19a), le stimulus est activé en continu. Dans la deuxième configuration (figure 4.19b), le stimulus est activé de manière répétée pendant la moitié du temps. Les trajectoires du haut (sans stimulus) et du milieu (sans sensibilisation) servent de référence. Un agent capable de sensibilisation devrait commencer par suivre la trajectoire du milieu, puis s'en éloigner progressivement en descendant davantage.

Ces graphiques ne sont pas les résultats de simulation mais sont "dessinés à la main" pour illustrer les propriétés de la sensibilisation spécifique au site.

### 4.2.1 Modèles

#### Modèle avec mémoire sensorielle

L'adaptation du modèle avec mémoire sensorielle concerne le calcul du senseur *sensibilisé*  $s_s$ , équivalent du senseur habitué  $s_h$  pour l'habituation, et dont la valeur est utilisée à la place de la valeur sensorielle brute  $s_b$  en tant que composante sensorielle de l'agent.

Nous utilisons le même calcul pour la mémoire sensorielle, rappelé ici :

$$\frac{d\delta}{dt}(\mathbf{x}_E, t) = \beta(s_b(\mathbf{x}_E) - \delta(\mathbf{x}_E, t)) \quad (4.17)$$

avec  $\beta = 0.25$ , paramètre qui conserve les mêmes propriétés : il est compris entre 0 et 1 et règle la sensibilité du modèle aux variations du stimulus. Plus il est élevé, plus l'est la sensibilité et plus rapidement  $\delta$  tend vers  $s_b$ .

La valeur du senseur *sensibilisé*  $s_s$  est calculée à partir de la valeur sensorielle brute  $s_b$  et de la mémoire sensorielle  $\delta$  :

$$s_s(\mathbf{x}_E, t) = s_b(\mathbf{x}_E)(1 + \gamma(1 - f_a(\mathbf{x}_E, t))) \quad (4.18)$$

$$f_a(\mathbf{x}_E, t) = \begin{cases} \frac{1}{1 + e^a} & \text{si } s_b(\mathbf{x}_E) > \epsilon \\ 0 & \text{sinon} \end{cases} \quad (4.19)$$

$$a = -10 \left( \frac{\|\delta(\mathbf{x}_E, t) - s_b(\mathbf{x}_E)\|_2}{s_b(\mathbf{x}_E)} - \frac{1}{2} \right) \quad (4.20)$$

avec  $\gamma = 0.25$ . Le calcul du facteur d'apprentissage  $f_a$  est inchangé, il est remplacé ici pour rappel et l'on pourra se référer à son tracé en figure 4.2. Le stimulus est perçu avec peu ou pas de changement si sa valeur est loin de la mémoire sensorielle ( $\|\delta - s_b\|$  proche de  $s_b$ ) et il est perçu comme plus intense si sa valeur est proche de la mémoire du capteur ( $\delta \approx s_b$ ), avec  $(1 + \gamma(1 - f_a(\mathbf{x}_E, t))) \in [1, 1 + \gamma]$ . Le paramètre  $\gamma$  permet de réguler cette intensification.

#### Modèles avec bonus sur les nœuds

Pour les modèles suivants, nous utilisons exactement le même concept de surprise que pour l'habituation et sa même définition. Par conséquent, les surprises sont toujours les écarts dans la perception du stimulus. Le changement se trouve au niveau de l'impact de ces événements surprenants : plutôt que d'affaiblir les nœuds proches de la position sensorimotrice de l'agent au moment de la surprise, nous les renforçons.

**Impact sur l'énergie des nœuds** L'énergie d'un nœud  $N$  reste  $N_e \in [0, 1]$ .

Contrairement à l'habituation, ici plus l'énergie est faible, plus l'influence du nœud sur la dynamique est *forte*. La surprise *sur* diminue l'énergie des nœuds qui est récupérée progressivement dans le temps. Nous conservons les mêmes équations pour la gestion de l'énergie, qui dépendent à la fois de la surprise *sur* et de la proximité entre le nœud  $N$  et la position sensorimotrice  $\mathbf{x}_{SM}$  de l'agent :

$$c_e(N, t) = ((1 + sur(t))^{\beta_e} - 1) \cdot \xi_N(\mathbf{x}_{SM}) \quad (4.21)$$

$$N_e(t + \Delta t) = N_e(t) + \gamma_e \Delta t - c_e(N, t) \quad (4.22)$$

avec  $\beta_e = 2$  régule la consommation d'énergie et  $\gamma_e = 0.05$  régule la récupération d'énergie.

La différence avec l'habitude se situe dans l'utilisation de l'énergie dans les vecteurs de vitesse et d'attraction :

$$\mathbf{V}(\mathbf{x}_{SM}) = \sum_{\mathcal{A}} \xi_N(\mathbf{x}_{SM}) \cdot \frac{1}{N_e + \epsilon} \cdot [\mathbf{N}_v]^\mu \quad (4.23)$$

$$\mathbf{A}(\mathbf{x}_{SM}) = \sum_{\mathcal{A}} \xi_N(\mathbf{x}_{SM}) \cdot \frac{1}{N_e + \epsilon} \cdot [\Gamma(\mathbf{N}_p - \mathbf{x}_{SM}, \mathbf{N}_v)]^\mu \quad (4.24)$$

Ce changement permet d'augmenter rapidement l'influence d'un nœud en fonction de l'énergie de façon inverse où, pour l'habitude, le modèle pouvait rendre cette influence nulle.

**Impact sur le poids des nœuds** Le modèle avec impact sur le poids des nœuds est plus difficile à modifier pour permettre de la sensibilisation spécifique au site. La cause est au moins, comme discuté dans la section 3, que l'impact du poids se traduit via le facteur  $\omega(N_w)$  (voir équation 3.3 et figure 3.3). En pratique, avoir un poids élevé n'impacte pas significativement la dynamique du système, en tout cas pas dans une configuration aussi simple que celles utilisées dans nos expériences.

Les changements apportés par rapport au modèle sur l'habitude sont un renforcement du poids en fonction de la surprise et de la proximité entre le nœud et la position sensorimotrice de l'agent mais sans être fonction du facteur  $\omega(N_w)$ . De plus, les valeurs des poids sont limitées dans  $[-2000, 2000]$ . Les équations utilisées sont :

$$g_w(N, t) = \beta_w \cdot sur(t) \cdot d(\mathbf{N}_p, \mathbf{x}_{SM}) \quad (4.25)$$

$$N_w(t + \Delta t) = N_w(t) + g_w(N, t) - \gamma_w \Delta t \quad (4.26)$$

avec  $\beta_w = 4000$  régulant le gain de poids et  $\gamma_w = 75$  pour la récupération spontanée.

Le choix d'avoir des poids négatifs et de ne pas prendre en compte  $\omega(N_w)$  pour le renfort du poids permet de ne pas entraver le renforcement des nœuds à poids faible.

#### 4.2.2 Expériences et résultats

$\Delta t$	$k_t$	$k_d$	$k_\omega$	$\tau$	$\mu$
0.1	1	30	0.025	10	$\in [0; 1]$

TABLE 4.3 – Valeurs des paramètres et options utilisées pour les résultats sur la sensibilisation spécifique au site (section 4.2.2)

#### Instanciation du modèle

Pour nos expériences, nous reprenons un IDSM similaire à celui pour l'habitude, avec deux nœuds comme indiqué sur la figure 4.20. La différence est que nous réduisons le  $\mathbf{N}_v$  des deux nœuds pour pouvoir amplifier de manière plus visible leur influence lors des expériences :

- $N^1 = \langle \mathbf{p} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -0.1 \\ 0 \end{bmatrix}, w = 0 \rangle,$
- $N^2 = \langle \mathbf{p} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix}, w = 0 \rangle.$

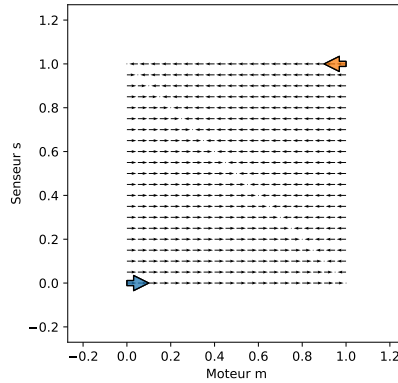


FIGURE 4.20 – IDSM avec deux nœuds utilisé dans les expériences sur la sensibilisation spécifique au site. Les flèches colorées représentent les nœuds, les flèches noires indiquent la dynamique motrice : la composante motrice  $[\mathbf{x}_{SM}]^\mu$  tend vers 0 lorsque le stimulus est perçu ou vers 1 lorsque le stimulus n'est pas perçu. Les comportements liés au moteur sont que l'agent descend plus ou moins vite lorsque la composante motrice  $[\mathbf{x}_{SM}]^\mu$  est dans  $[0; 0.5]$ , l'agent est à l'arrêt lorsque  $[\mathbf{x}_{SM}]^\mu$  vaut 0.5 et l'agent monte plus ou moins vite lorsque  $[\mathbf{x}_{SM}]^\mu$  est dans  $]0.5, 1]$ .

Les comportements résultants sont similaires à ceux de l'habituation, à part que l'agent accélère et décélère plus progressivement. Si l'agent perçoit le stimulus (la valeur sensorielle brute est  $s = 0.75$ ), sa vitesse diminue (le moteur tend vers  $-1$ ) et il va se mettre à descendre ; si l'agent ne perçoit pas le stimulus (la réponse sensorielle est  $s = 0$ ), sa vitesse augmente (le moteur tend vers  $1$ ) et il va se mettre à monter. La valeur sensorielle brute lorsque le stimulus est activé est  $0.75$  plutôt que  $1$  pour permettre une sensibilisation avec la mémoire sensorielle.

### Configuration expérimentale pour la sensibilisation spécifique au site

Afin de mesurer si nos modèles sont capables de sensibilisation spécifique au site, nous utilisons le protocole suivant.

Comme pour l'habituation, l'expérience est divisée en des phases de stimulation (avec un stimulus continu ou répété) et des phases de calme (sans aucun stimulus). Nous conservons l'alternance des phases : deux phases de stimulation pour évaluer la sensibilisation, puis une phase de calme suivie d'une phase de stimulation pour évaluer la récupération spontanée de la sensibilisation. Chaque phase dure vingt unités de temps. Pour limiter les différences entre les conditions expérimentales, au début de chaque phase, l'agent est réinitialisé à une position de départ à vitesse nulle ( $\mathbf{x}_E = \dot{\mathbf{x}}_E = \ddot{\mathbf{x}}_E = 0$ ) et à un état sensorimoteur neutre  $\mathbf{x}_{SM} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$ .

Nous conservons aussi les schémas de stimulation :

- (a) : stimulus continu ;
- (b) : répétitions de 3 unités de temps avec stimulus / 1 unité de temps sans ;
- (c) : répétitions de 2 unités de temps avec / 2 unités de temps sans ;
- (d) : répétitions de 1 unité de temps avec / 3 unités de temps sans.



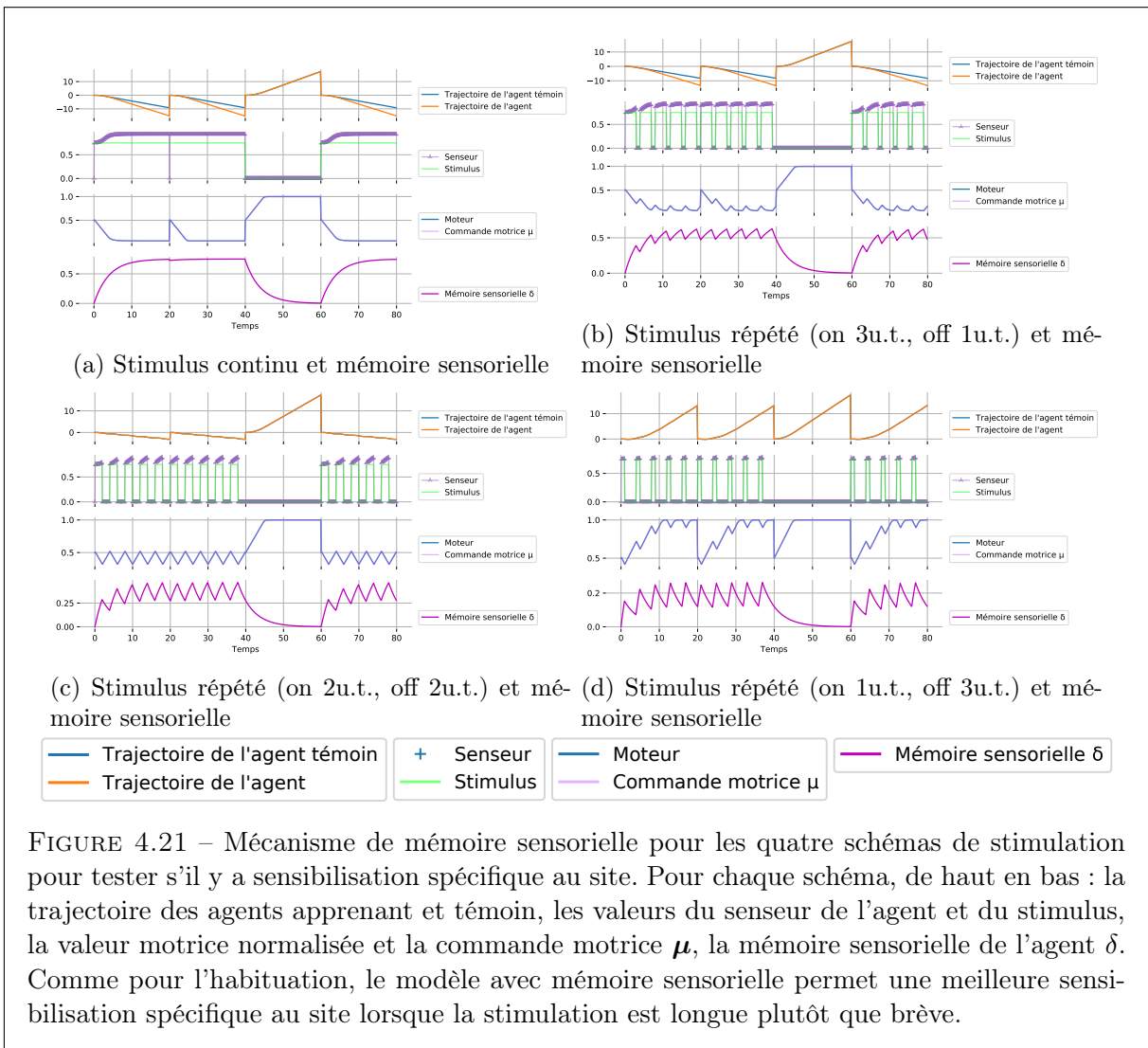


FIGURE 4.21 – Mécanisme de mémoire sensorielle pour les quatre schémas de stimulation pour tester s'il y a sensibilisation spécifique au site. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du capteur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$ . Comme pour l'habituation, le modèle avec mémoire sensorielle permet une meilleure sensibilisation spécifique au site lorsque la stimulation est longue plutôt que brève.

## Résultats pour la sensibilisation spécifique au site

La sensibilisation spécifique au site est évaluée en comparant le comportement d'un agent apprenant (utilisant un de nos modèles) avec celui d'un agent témoin (sans mécanisme d'apprentissage). Les résultats pour chaque modèle sont donnés sur les figures 4.21-4.23.

Pour le modèle avec mémoire sensorielle (voir figure 4.21), plus le stimulus reste activé et plus la mémoire sensorielle se rapproche de la valeur sensorielle brute, ce qui résulte en une meilleure sensibilisation. La récupération n'est pas visible à travers le comportement de l'agent qui descend globalement toujours la même distance entre phases 1, 2 et 4, mais elle est bien observable à travers la valeur du capteur (deuxième graphique) et la valeur de la mémoire sensorielle  $\delta$  (quatrième graphique).

Pour les modèles avec impact sur les nœuds, l'interprétation des résultats est plus compliquée. Pour l'habituation, un nœud affaiblit influence moins l'agent qui va alors mettre moins de distance entre lui et le nœud, puis plus l'agent est proche du nœud à chaque surprise, plus le nœud est

à nouveau affaiblit (voir figure 4.7). Cela permet à terme de stabiliser sur une même valeur motrice et conduit soit à de l'habituation soit à de l'habituation inverse. Avec la sensibilisation et cette configuration, l'état est moins stable : un nœud repousse l'agent plus loin de lui et plus proche de l'autre nœud (sur la dimension motrice). Par conséquent, cela favorise à terme le renfort des deux nœuds plutôt que celui d'un seul nœud. Avec les deux modèles, comme pour l'habituation, le manque de surprise avec le schéma **(a)** ne permet pas de sensibilisation. Le modèle avec énergie (voir figure 4.22) montre de la sensibilisation *inverse* marquée pour le schéma **(b)** et très légère pour le schéma **(c)**. Avec le schéma **(d)**, l'agent monte moins que l'agent témoin, il y a bien sensibilisation. En poussant un peu plus profondément l'analyse, nous pouvons voir un phénomène intéressant : l'énergie augmente tellement l'influence des deux nœuds que la trajectoire motrice devient de plus en plus brusque, ce qui se traduit au niveau comportemental par un agent plus réactif aux activations et désactivations du stimulus. La modification des  $N_v$  à 0.1 rend les agents plus lents à changer de comportements par rapport aux expériences sur l'habituation. Cela est visible autant sur les trajectoires motrices que sur les trajectoires de l'agent témoin par exemple : les "vaguelettes" d'activation et désactivation du stimulus lors des schémas **(b)**, **(c)** et **(d)** ne sont pas visibles sur la trajectoire de l'agent témoin (comparer entre les figures 4.5 pour l'habituation et 4.22 pour la sensibilisation). Au contraire, avec l'agent basé sur le modèle avec énergie ces "vaguelettes" redeviennent très marquées, avec une position motrice allant directement à 1 lors de l'activation ou à 0 lors de la désactivation du stimulus (voir figure 4.22d).

Concernant le modèle avec bonus sur les poids (figure 4.23), nous n'observons aucun changement en schéma **(a)** par manque de surprise, une sensibilisation inverse pour les schémas **(b)** et **(c)** et une sensibilisation avec le schéma **(d)**. L'évolution des poids est en accord avec cette observation : malgré globalement une baisse constante, le poids du nœud  $N^2$  est supérieur à celui de  $N^1$  pendant les schémas **(b)** et **(c)**, favorisant le comportement montant ; inversement pour le schéma **(d)**. Cependant, contrairement au modèle avec l'énergie, ici les nœuds n'arrivent à égalité que lorsque leurs poids sont tous deux vers  $-2000$  (la borne inférieure des valeurs des poids), cela suite à l'affaiblissement continu des poids dans le temps (voir l'équation 4.26). Le comportement résultant est donc un agent globalement plus lent que l'agent témoin, avec une trajectoire motrice restant plus près de 0.5 (normalisée, donc moteur à l'arrêt), et ce malgré un paramètre  $\beta_w = 4000$  (régulant le renforcement des nœuds) huit fois plus important que son équivalent en habituation. La cause est le fait d'avoir la valeur sensorielle brute lorsque le stimulus est activé à 0.75 plutôt que 1. Cela apporte un biais : sans modification d'influence des nœuds, l'agent va de  $[0.25, 0]$  à  $[1, 0]$  hors stimulation puis de  $[1, 0.75]$  à  $[0.25, 0.75]$  lors de la stimulation. Avant apprentissage, l'agent est donc au plus près de 0.25 de chaque nœud. Suite à cette baisse régulière du poids, la récupération de la sensibilisation (inverse ou non) n'opère pas vraiment car pour comparer il faudrait alors initialiser les poids à  $-2000$ .

En conclusion de cette section, nous constatons que les trois modèles proposés présentent une sensibilisation spécifique au site. Pour la récupération, les résultats sont plus mitigés, même si nous pouvons observer qu'elle a bien lieu au niveau des mécanismes pour la mémoire sensorielle et pour l'énergie. Comme pour l'habituation, le rapport de stimulation a un impact important sur l'amplitude de la sensibilisation, impact qui varie selon le modèle utilisé. L'activation continue ne déclenche la sensibilisation qu'avec le modèle avec mémoire sensorielle. Un temps d'activation du stimulus supérieur ou égale au temps de désactivation provoque une sensibilisation *inverse* pour les modèles avec impact sur les nœuds, où le comportement renforcé est le comportement hors stimulation (l'agent monte plus rapidement). Un autre phénomène notable est la sensibilisation

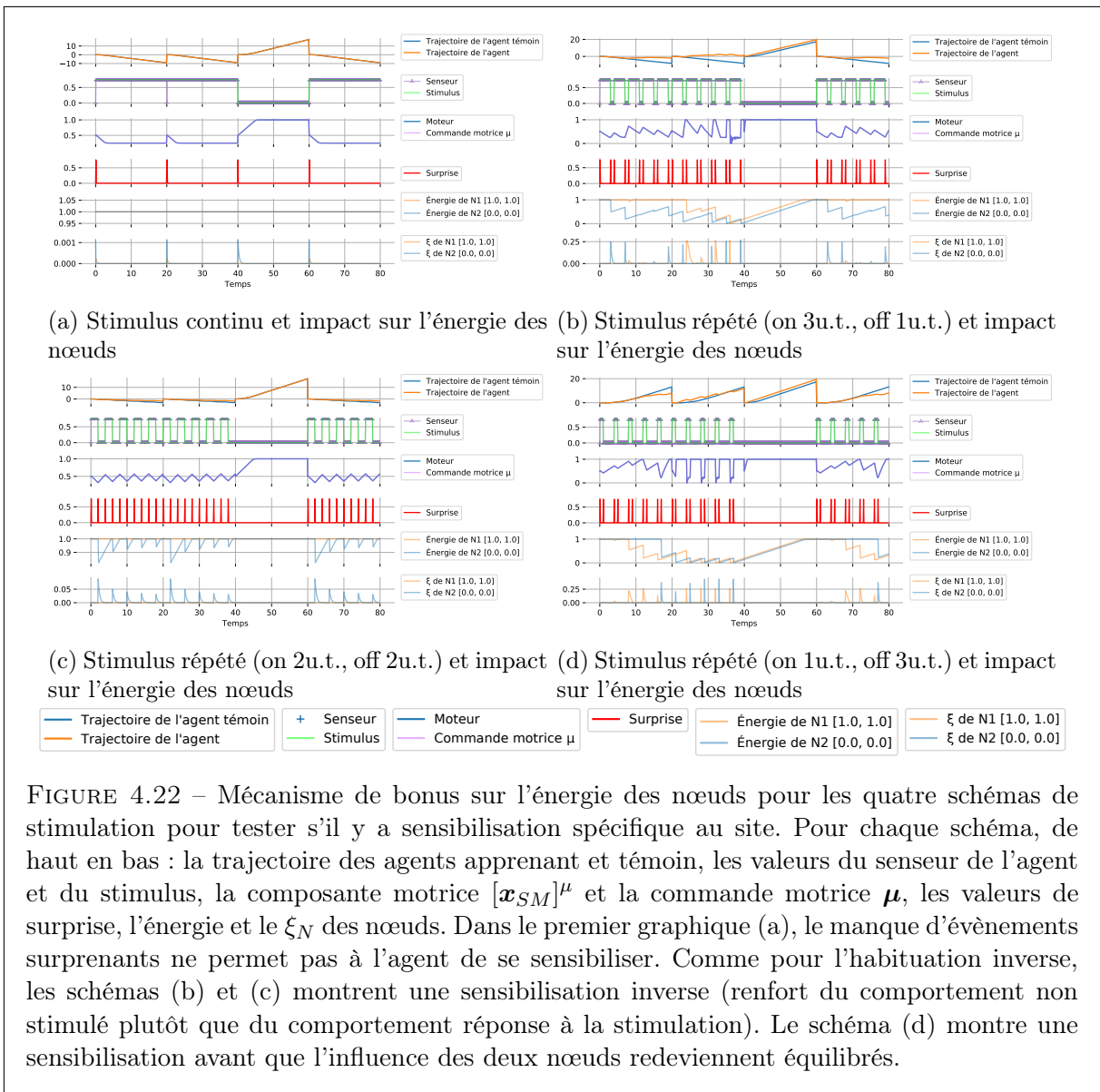
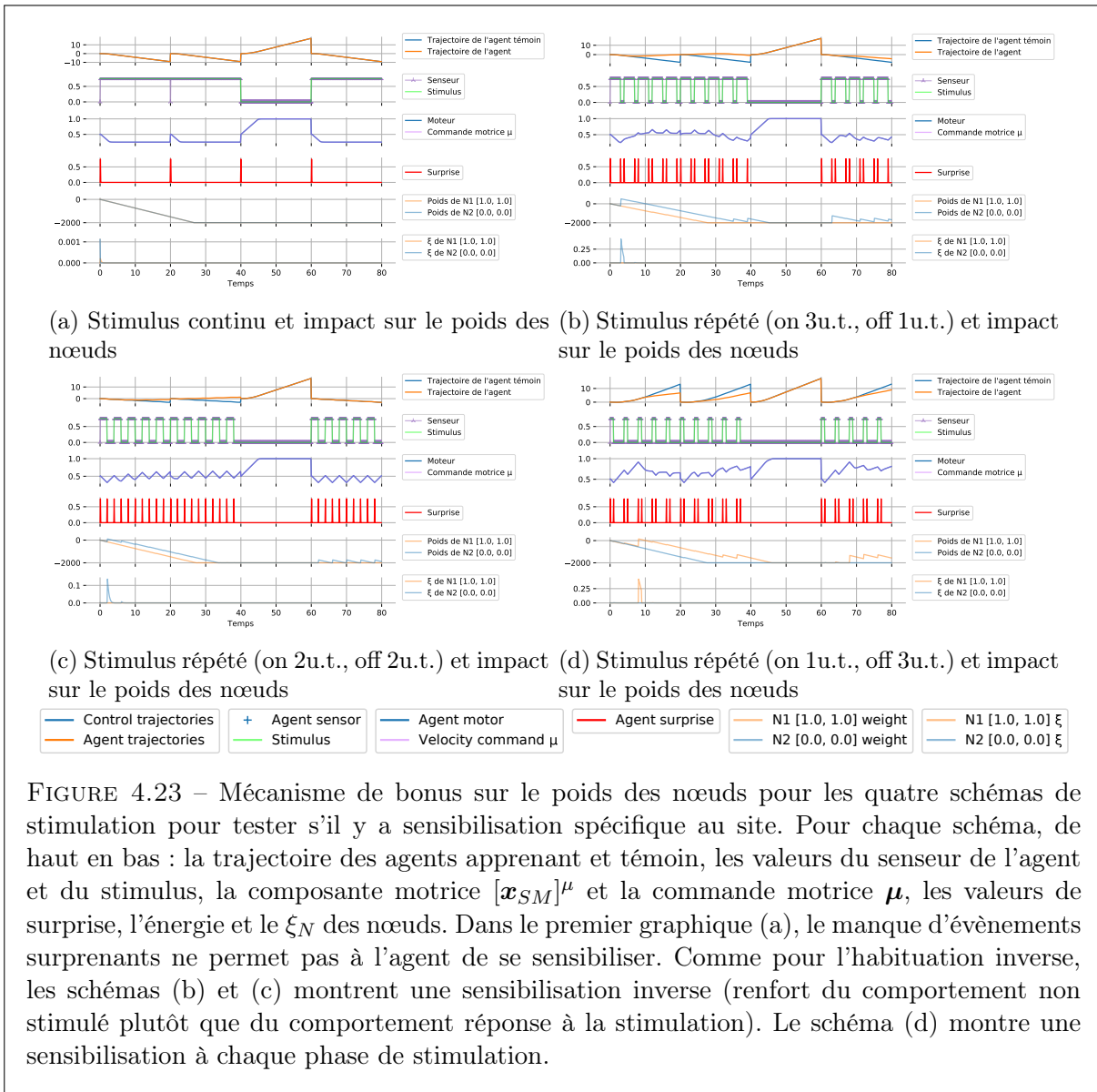


FIGURE 4.22 – Mécanisme de bonus sur l'énergie des nœuds pour les quatre schémas de stimulation pour tester s'il y a sensibilisation spécifique au site. Pour chaque schéma, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la composante motrice  $[\mathbf{x}_{SM}]^\mu$  et la commande motrice  $\mu$ , les valeurs de surprise, l'énergie et le  $\xi_N$  des nœuds. Dans le premier graphique (a), le manque d'événements surprenants ne permet pas à l'agent de se sensibiliser. Comme pour l'habituation inverse, les schémas (b) et (c) montrent une sensibilisation inverse (renfort du comportement non stimulé plutôt que du comportement réponse à la stimulation). Le schéma (d) montre une sensibilisation avant que l'influence des deux nœuds redeviennent équilibrés.



à l'activation et à la désactivation observée avec le modèle impactant l'énergie des nœuds (voir figure 4.22d), phénomène dû à un renfort important des deux nœuds.

### Combinaison des modèles

Suite aux résultats positifs observés pour les combinaisons pour l'habituation pour l'observer pour tous les schémas de stimulation, nous combinons nos modèles avec l'objectif de pouvoir maintenant observer de la sensibilisation spécifique au site pour l'ensemble des schémas à partir d'une même combinaison. Nous expérimentons en combinant les modèles avec mémoire sensorielle et impact sur l'énergie des nœuds, puis mémoire sensorielle et impact sur le poids des nœuds.

Les résultats pour la combinaison mémoire sensorielle et énergie sont visibles sur la figure 4.24. Contrairement à l'habituation, la sensibilisation inverse provoquée par le modèle avec énergie ne devient pas une sensibilisation "normale". L'agent du schéma **(a)** a le même comportement (sensibilisation) qu'avec le modèle avec mémoire sensorielle uniquement, celui du schéma **(c)** a le même comportement qu'avec le modèle avec énergie uniquement (faible sensibilisation inverse). Pour les schémas **(b)** et **(d)**, la mémoire sensorielle permet à l'agent d'avoir une composante sensorielle plus proche de 1 (surtout dans le cas du schéma **(b)**), et par conséquent cela augmente l'effet de l'impact de la surprise, renforçant les deux nœuds plus fortement plus rapidement. C'est pourquoi la trajectoire motrice alterne plus brusquement entre 0 et 1, et que l'agent devient plus sensible aux changements du stimulus (activation ou désactivation).

Pour la combinaison mémoire sensorielle et bonus sur les poids, la figure 4.25 ne montre que le schéma **(b)**. Pour les trois autres schémas, les résultats sont identiques ou similaires aux résultats avec modèles individuels (mémoire sensorielle pour le schéma **(a)**, modèle avec les poids pour les schémas **(c)** et **(d)**). Nous pouvons voir sur la figure que, comme pour la combinaison mémoire sensorielle et énergie, la mémoire sensorielle permet à l'agent d'avoir une composante sensorielle plus proche de 1, augmentant l'effet de l'impact de la surprise sur les poids, qui sont plus fortement renforcés qu'en absence du modèle avec mémoire sensorielle (voir en comparaison la figure 4.23b). La trajectoire motrice des phases 2 et 4 reflète de manière marquée les modifications dans la stimulation (descendante quand le stimulus est actif, montante sinon). Contrairement au modèle avec énergie, la composante  $N_v$  des nœuds n'est pas renforcée et par conséquent le comportement de l'agent est moins impacté : il monte légèrement en début de phase et descend légèrement en fin de phase, sans qu'il y ait de lien direct visible avec la stimulation. Enfin, comme pour le modèle avec bonus sur les poids seul, la récupération de la sensibilisation n'a pas lieu car les poids sont trop différents de 0 en début de phase 4 (la dynamique en début de phase est proche de celle illustrée sur la figure 4.14d).

Les résultats de la combinaison de modèles pour la sensibilisation spécifique au site sont moins marqués que ceux pour l'habituation. En particulier, nous n'observons pas de synergie entre nos modèles permettant un changement de la sensibilisation inverse des modèles avec surprise vers une sensibilisation "normale". Cependant, la variation de la valeur du senseur due à la mémoire sensorielle est aussi plus restreinte que celle qui a lieu lors des expériences sur l'habituation : ici, la variation permet au senseur d'aller de 0.75 à 1 là où pour l'habituation il peut aller de 1 à 0. La surprise est d'autant moins impactée par cette variation, ce qui explique au moins en partie la plus forte ressemblance entre les résultats des modèles seuls et ceux des modèles combinés pour la sensibilisation spécifique au site. Le schéma **(b)** est celui permettant aux modèles avec surprises de bénéficier le plus de la mémoire sensorielle. En effet, il s'agit du schéma avec répétition du stimulus (donc plusieurs activations et désactivations permettant autant événements de surprises)

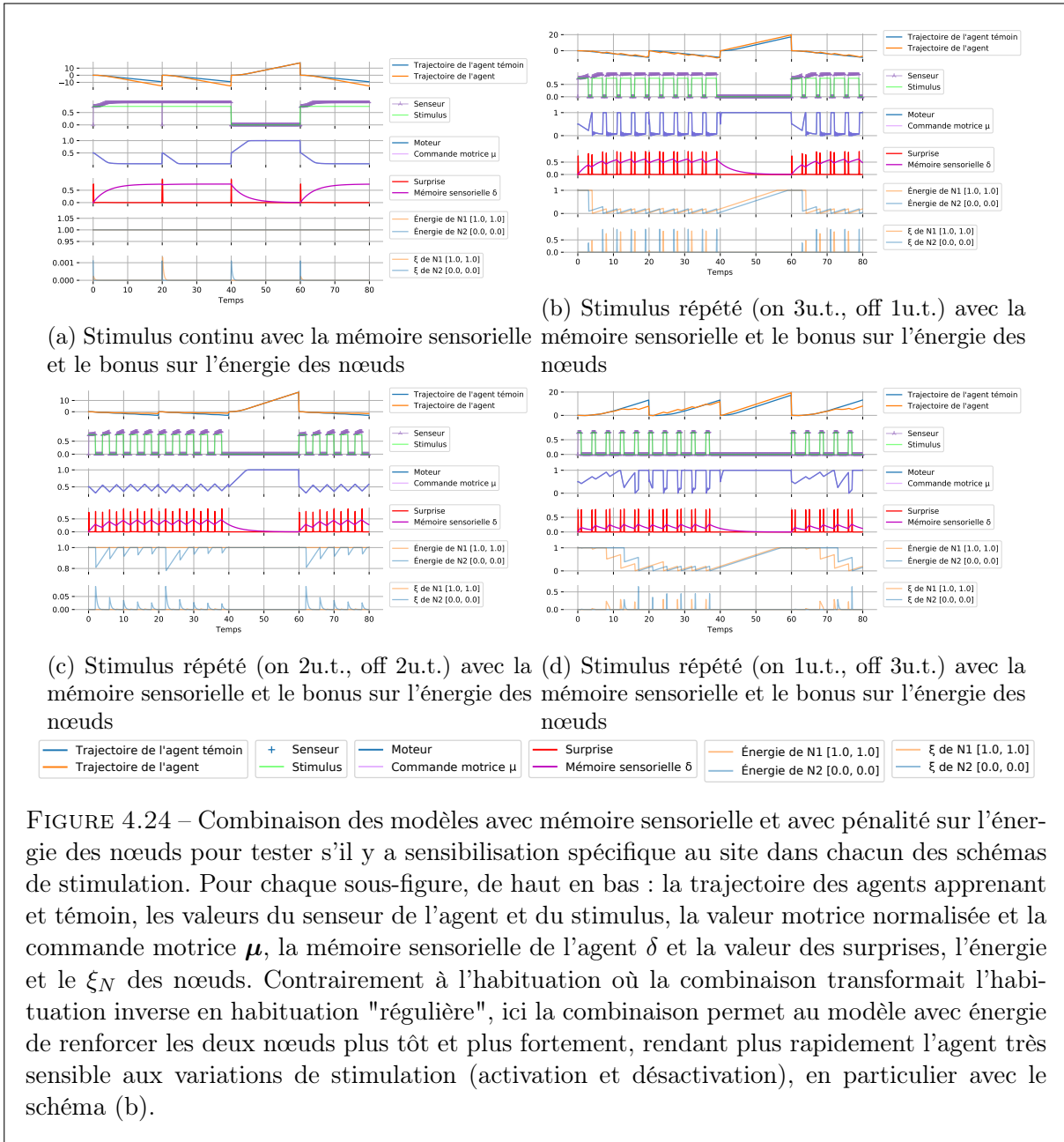
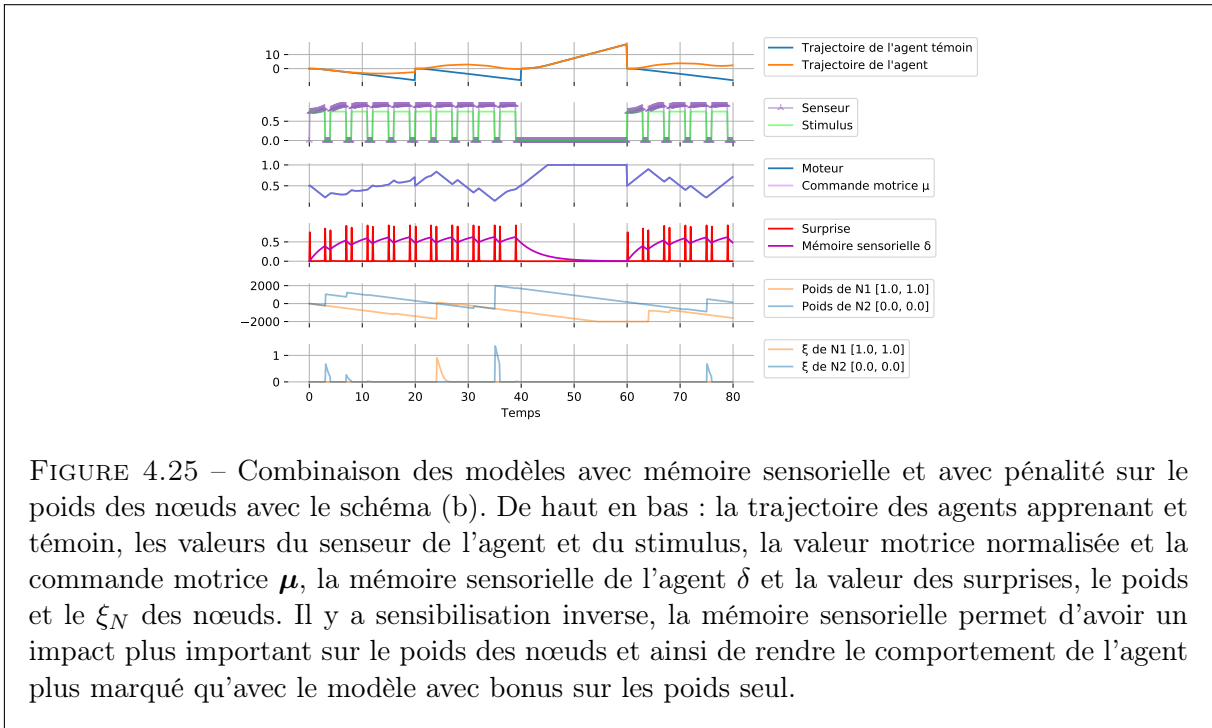


FIGURE 4.24 – Combinaison des modèles avec mémoire sensorielle et avec pénalité sur l'énergie des nœuds pour tester s'il y a sensibilisation spécifique au site dans chacun des schémas de stimulation. Pour chaque sous-figure, de haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. Contrairement à l'habituation où la combinaison transformait l'habituation inverse en habituation "régulière", ici la combinaison permet au modèle avec énergie de renforcer les deux nœuds plus tôt et plus fortement, rendant plus rapidement l'agent très sensible aux variations de stimulation (activation et désactivation), en particulier avec le schéma (b).



avec le temps d’activation le plus long (et par conséquent une mémoire sensorielle plus élevée, augmentant la valeur de la surprise).

### En utilisant l’IDSM déformable

Pour clore nos expériences sur la sensibilisation spécifique au site, nous effectuons des tests préliminaires afin d’observer s’il y a une compatibilité possible entre nos modèles de sensibilisation et l’IDSM pleinement déformable. Pour ces expériences, nous activons la création de nœuds et la modification continue des poids des nœuds.

Une première observation est que la trajectoire de l’agent témoin change de manière plus ou moins visible peu après 10 unités de temps. Cela est dû à l’activation des nouveaux nœuds avec  $\tau = 10$ .

Le modèle avec mémoire sensorielle (voir figure 4.26) permet toujours d’exprimer de la sensibilisation spécifique au site avec le schéma (a). L’écart entre les phases est très léger, comme lors des expériences avec l’IDSM simplifié. Nous pouvons voir par contre que l’agent témoin descend de plus en plus entre la phase 1, 2 et 4.

Le modèle avec bonus sur l’énergie lors du schéma (d) (voir figure 4.27b) semble marcher "mieux" qu’en IDSM simplifié, surtout en phase 4 où l’agent est en permanence moins haut que l’agent témoin. Cependant, il faut prendre en compte qu’avec les nouveaux nœuds et l’instabilité propre à la sensibilisation dans cette configuration, les phases 1, 2 et 4 sont très différentes les unes des autres, surtout si l’on regarde la trajectoire motrice qui ne semble pas se stabiliser sur un motif clair. Dans ces conditions, il devient compliqué d’évaluer la récupération de la sensibilisation à part en se référant aux valeurs de l’énergie des nœuds initiaux, mais il faudrait aussi regarder celle des autres nœuds.

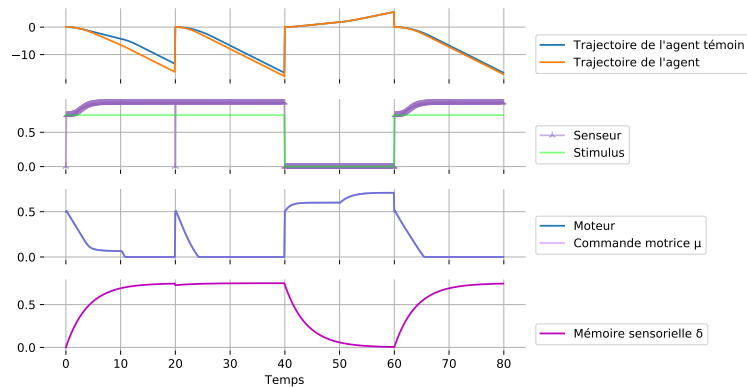


FIGURE 4.26 – Modèle avec la mémoire sensorielle basé sur l'IDSM original lors du schéma (a) pour tester s'il y a sensibilisation spécifique au site. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$ . La sensibilisation spécifique au site est toujours observable même si de nouveaux nœuds sont créés et les poids des nœuds continuellement mis à jour.

Le comportement de l'agent avec le schéma (c) (figure 4.27a) diverge de manière importante de celui de l'agent témoin, il montre globalement une sensibilisation inverse. Cette sensibilisation est de plus en plus marquée selon les phases de stimulations (phases 1, 2 et 4) et est aussi très marquée pendant la phase de calme (phase 3). L'évolution de la trajectoire motrice montre qu'avec l'augmentation d'énergie et la création de nœuds l'agent peut être amené à évoluer selon une dynamique très instable, avec des périodes plus ou moins longues où la composante motrice  $[\mathbf{x}_{SM}]^\mu$  oscille très rapidement entre 0 et 1. Cette instabilité perdure un peu en début de phase de calme (phase 3), mais la dynamique explorée ensuite est au contraire assez rigide. Il n'est pas clair pour nous pour le moment de savoir si l'agent pourrait retourner dans une telle dynamique mouvementée si l'expérience durait plus longtemps, mais c'est un phénomène à prendre en compte si l'amplification de l'influence des nœuds avec l'énergie tend vers l'"infini" comme actuellement. Pour éviter un tel phénomène, une solution serait d'ajouter un paramètre d'énergie minimale permettant de maintenir l'énergie à une valeur non nulle et plus globalement de paramétrer l'effet maximum de l'impact à travers l'énergie.

Pour la combinaison des modèles avec mémoire sensorielle et bonus sur les poids, les résultats sont similaires à ceux avec l'IDSM simplifié : sensibilisation pour le modèle avec mémoire sensorielle, sensibilisation inverse et sensibilité à l'activation et désactivation du stimulus pour le modèle avec énergie, et renfort de ce dernier phénomène lorsque les modèles sont combinés. Il y a cependant des particularités qui sont toutes illustrés sur la figure 4.28.

Le modèle avec mémoire sensorielle (figure 4.28a) montre un cas de sensibilisation inverse lors de la phase de calme : il monte plus haut que le témoin. Ce cas particulier s'explique avec l'activation des nouveaux nœuds créés en début de phase de calme, nœuds incitant à monter. Dans le cas de l'agent avec mémoire sensorielle, ces nœuds sont plus influents que ceux de l'agent témoin car les nouveaux nœuds créés auparavant dans la zone liée à l'activation du stimulus sont plus éloignés. En effet, pour l'agent témoin, ces nœuds ont tous une position sensorielle de



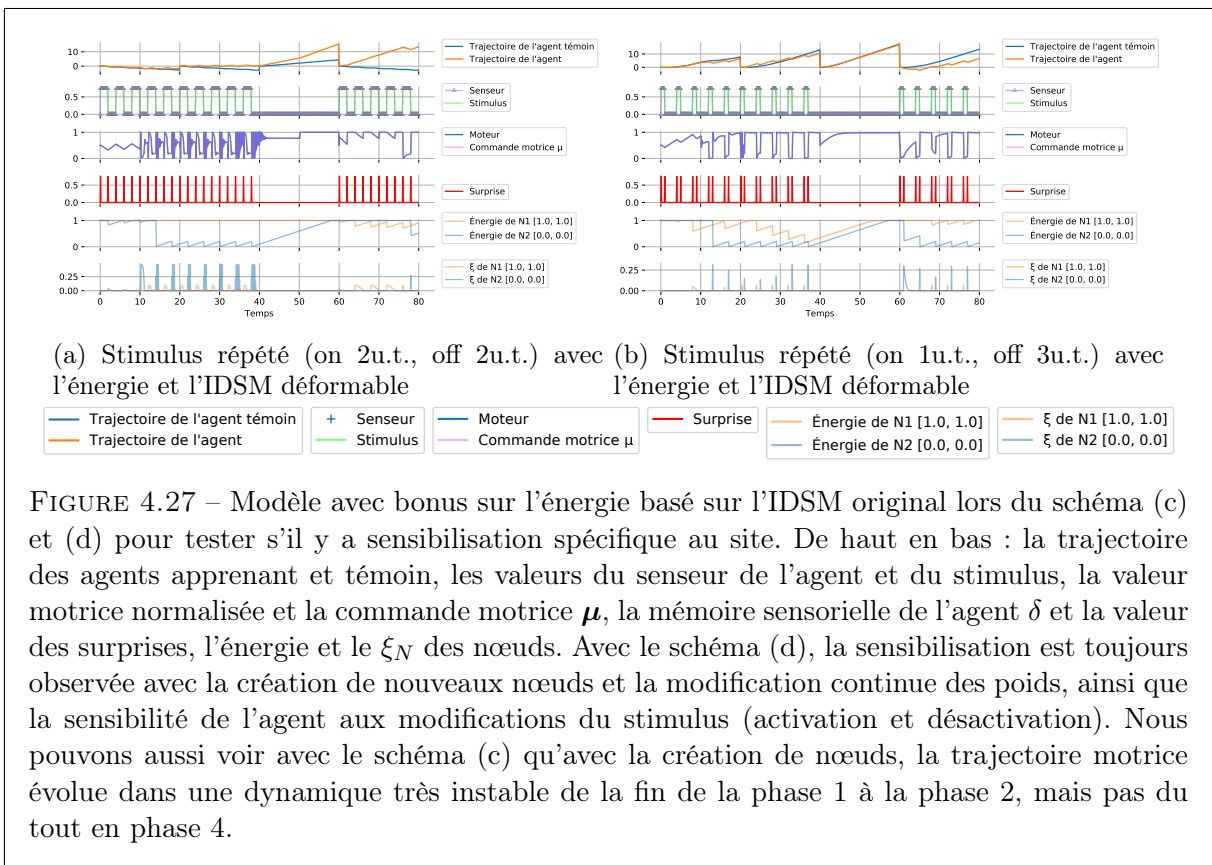


FIGURE 4.27 – Modèle avec bonus sur l'énergie basé sur l'IDSM original lors du schéma (c) et (d) pour tester s'il y a sensibilisation spécifique au site. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du capteur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. Avec le schéma (d), la sensibilisation est toujours observée avec la création de nouveaux nœuds et la modification continue des poids, ainsi que la sensibilité de l'agent aux modifications du stimulus (activation et désactivation). Nous pouvons aussi voir avec le schéma (c) qu'avec la création de nœuds, la trajectoire motrice évolue dans une dynamique très instable de la fin de la phase 1 à la phase 2, mais pas du tout en phase 4.

0.75 (valeur sensorielle brute) là où ceux de l'agent avec mémoire sensorielle sont "étalés" dans  $[0.75, 1]$ .

Pour le modèle avec énergie (figure 4.28b), la création de nouveaux nœuds permet d'avoir plus rapidement un équilibre entre les nœuds qu'avec l'IDSM simplifié. Cela se traduit par une composante motrice passant d'un extrême à l'autre suite aux variations (activation et désactivation) du stimulus plus tôt par rapport au même modèle et schéma sans la création de nœuds (voir figure 4.22b). La combinaison des deux modèles (figure 4.28c) stabilise et rigidifie encore plus rapidement ce phénomène, la création de nouveaux nœuds supprime les oscillations qui existent dans le même cas avec l'IDSM simplifié (voir figure 4.24b).

Le modèle avec bonus sur les poids nécessite plus de travail pour régler sa compatibilité avec l'IDSM original, comme pour l'habituation aucun test effectué pour le moment ne s'est révélé concluant.

### 4.2.3 Bilan sur la sensibilisation spécifique au site

Les modèles initialement pensés pour l'habituation ont pu être assez facilement modifiés pour répondre à la question de la sensibilisation spécifique au site, du fait que ces deux apprentissages produisent des effets comportementaux inverses l'un de l'autre.

Nos trois modèles sont capables de permettre à l'agent de se sensibiliser au stimulus perçu, la mémoire sensorielle lorsque le stimulus est activé plus longtemps que désactivé, les modèles avec impact sur les nœuds lorsque le stimulus est bref. De plus, ces modèles avec impact sur les nœuds montrent aussi une sensibilisation inverse où l'agent exprime plus fortement son comportement ascendant, comportement normalement exprimé hors stimulation.

Avec notre configuration, pour l'impact sur les nœuds, la sensibilisation est divergente là où l'habituation est convergente : un nœud renforcé éloigne l'agent de sa position  $N_p$ . Cela provoque un cas particulier : lorsque le modèle avec énergie renforce fortement les deux nœuds, l'agent devient très sensible aux variations du stimulus (activation ou désactivation).

Les remarques faites dans la section 4.1.3 sont toujours applicables sur ces adaptations. En particulier, le modèle avec mémoire sensorielle n'impacte toujours pas la dynamique de l'espace sensorimoteur directement mais la composante sensorielle de l'agent. Cela le "force" ainsi à avoir sa position sensorimotrice  $x_{SM}$  explorant une zone différente de l'espace sensorimoteur pour une valeur perçue du stimulus identique, contrairement aux modèles avec bonus sur les nœuds qui eux modifient la dynamique pour indirectement "déplacer" la position sensorimotrice de l'agent dans l'espace sensorimoteur. Les modèles impactant les nœuds ont toujours les mêmes différences fondamentales, celui sur l'énergie modifiant l'influence d'un nœuds sans modifier la distance où cette influence a lieu là où le modèle sur les poids change cette distance en ayant un impact moindre sur l'influence en elle-même (voir figure 4.14).

Les expériences sur les combinaisons de modèles sont moins concluantes que celles effectuées pour l'habituation. Cependant, les variations de la valeur du senseur provoquées par la mémoire sensorielle, même si elles sont moins importantes que pour l'habituation, restent suffisantes pour avoir un impact sur le mécanisme de surprise et ainsi améliorer les effets des modèles avec impact sur les nœuds.

Enfin, les premières expériences avec l'IDSM pleinement déformable montrent que nos modèles

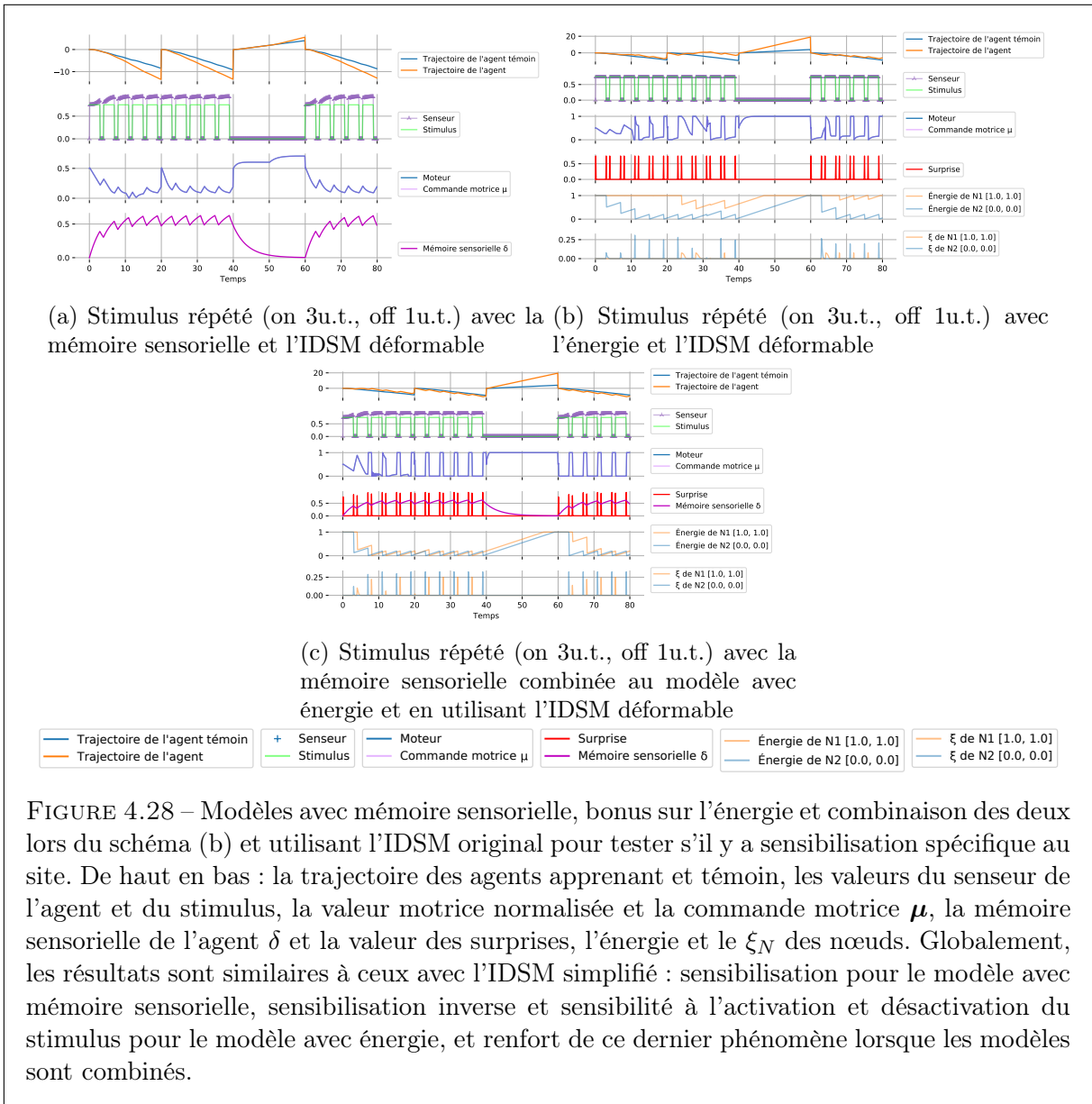


FIGURE 4.28 – Modèles avec mémoire sensorielle, bonus sur l'énergie et combinaison des deux lors du schéma (b) et utilisant l'IDSM original pour tester s'il y a sensibilisation spécifique au site. De haut en bas : la trajectoire des agents apprenant et témoin, les valeurs du senseur de l'agent et du stimulus, la valeur motrice normalisée et la commande motrice  $\mu$ , la mémoire sensorielle de l'agent  $\delta$  et la valeur des surprises, l'énergie et le  $\xi_N$  des nœuds. Globalement, les résultats sont similaires à ceux avec l'IDSM simplifié : sensibilisation pour le modèle avec mémoire sensorielle, sensibilisation inverse et sensibilité à l'activation et désactivation du stimulus pour le modèle avec énergie, et renfort de ce dernier phénomène lorsque les modèles sont combinés.

peuvent être utilisés et continuer d'exprimer de la sensibilisation spécifique au site malgré cette complexité supplémentaire. Mais comme pour l'habituation, cette compatibilité ne doit pas être considérée comme acquise sans davantage d'expériences et analyses plus poussées. En particulier, nous avons déjà pu voir que la caractéristique de création de nœuds peut à la fois provoquer des instabilités lorsque combinée avec l'énergie renforçant infiniment l'influence des nœuds ainsi que de la rigidité avec la suppression de l'effet de la récupération sur le comportement, alors qu'il n'y a pas ce problème dans les expériences sur l'habituation.

### 4.3 Pseudo-conditionnement

Le dernier modèle que nous proposons concerne le pseudo-conditionnement (voir section 2.1.2). Cette forme de sensibilisation permet à une réponse comportementale à un stimulus donné d'être exprimée comme réponse à un autre stimulus. En tant que tel, il s'agit d'une sorte de généralisation de cette réponse à d'autres stimuli. Le modèle proposé vise à étendre la dynamique de réponse d'une dimension sensorielle de l'IDSM à une autre. Nous équipons chaque nœuds d'un masque adaptatif qui détermine l'étendue de l'influence d'un nœud dans l'espace sensorimoteur. Bien qu'initialement spécifique à l'état sensorimoteur du nœud, le pseudo-conditionnement adapte le masque pour élargir l'influence de ce nœud. Cette adaptation permet à des états sensorimoteurs de devenir plus "sensibles" au nœud modifié qu'auparavant.

Pour ce dernier apprentissage, nous illustrons aussi le comportement que nous attendons de notre modèle à l'aide d'une expérience de pensée. Cette expérience, illustrée sur la figure 4.19, utilise une configuration légèrement plus complexe puisqu'il est nécessaire de faire intervenir deux stimuli. L'objectif est d'obtenir une réponse au stimulus conditionnel *SC* similaire à la réponse au stimulus inconditionnel *SI* après que l'agent n'aie été exposé qu'au *SI* pendant un laps de temps. L'agent dispose alors de deux senseurs (chacun sensible spécifiquement à l'un des stimuli) en plus de son unique moteur. Les stimuli dans l'environnement peuvent être activés ou désactivés, ils sont perceptibles par l'agent avec des valeurs constantes quelle que soit la position de l'agent dans l'environnement. L'agent peut se déplacer le long d'une seule dimension, vers le haut ou vers le bas. Son comportement réflexe est de monter en l'absence de stimulus et de descendre lorsque l'agent perçoit le stimulus inconditionnel *SI*. Le stimulus conditionnel *SC* est neutre et ne provoque pas de réponse particulière avant apprentissage : l'agent monte comme s'il ne percevait pas de stimulus. Dans un tel contexte, si le pseudo-conditionnement réussit, nous devrions observer une généralisation de la réponse au stimulus inconditionnel *SI* comme réponse au stimulus conditionnel *SC* : l'agent descend lorsqu'il perçoit le stimulus conditionnel *SC* à la place de monter.

Dans la figure 4.29, nous considérons deux configurations et, pour chacune, nous montrons comment l'agent est censé se comporter lorsqu'il perçoit le stimulus conditionnel *SC* après avoir été exposé au stimulus inconditionnel *SI*, dans le cas où il est incapable de pseudo-conditionnement (trajectoire du haut lors de la stimulation conditionnelle) et dans le cas où il en est capable (trajectoire du bas). Lorsqu'un stimulus est présenté, il peut être continu ou répété et, pour faciliter la lecture, les périodes où un stimulus est actif sont représentées par un fond coloré (sombre pour le stimulus inconditionnel et clair pour le stimulus conditionnel).

Dans la première configuration (figure 4.29a), le stimulus inconditionnel *SI* est activé en continu pendant 20 unités de temps avant d'être désactivé, puis le stimulus conditionnel *SC* est activé à son tour pendant 20 unités de temps. Dans la deuxième configuration (figure 4.29b),

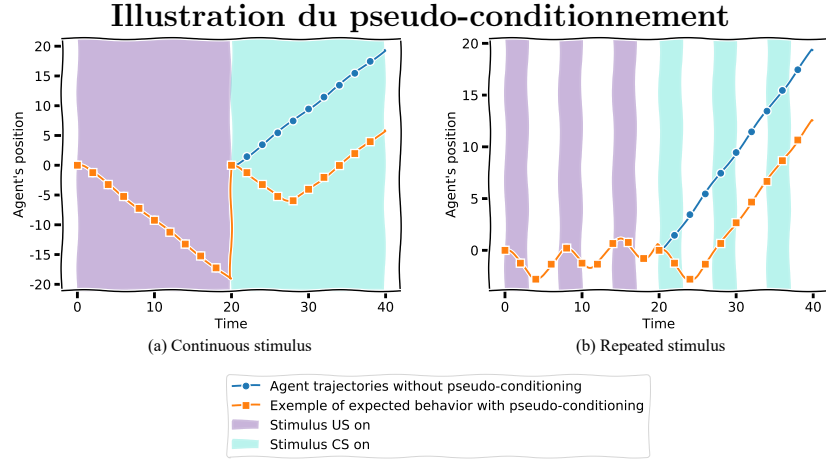


FIGURE 4.29 – Illustration des comportements attendus avec ou sans pseudo-conditionnement. Lorsque le stimulus inconditionnel  $SI$  ( $US$  sur la figure) est perçu, le comportement prédéfini de l'agent est de descendre. Hors stimulation, son comportement prédéfini est de monter. Lorsque le stimulus conditionnel  $SC$  ( $CS$  sur la figure) est perçu, ce stimulus étant neutre, il se comporte comme hors stimulation et monte. Avec le pseudo-conditionnement, nous nous attendons à ce qu'après une exposition à stimulus inconditionnel  $SI$ , l'agent apprenne à répondre de la même manière lorsqu'il perçoit le stimulus conditionnel  $SC$ , c'est-à-dire qu'il descende à la place de monter, puis qu'il se remette à monter après un certain temps sans exposition au stimulus inconditionnel  $SI$ . Les stimulations peuvent être continues (a) ou répétées (b).

les stimulations sont dans le même ordre mais elles sont de manière répétée plutôt qu'en continu. Lors de la phase avec stimulus inconditionnel, il ne devrait pas y avoir de différences entre un agent capable de pseudo-conditionnement et un agent qui en serait incapable. Lors de la phase qui suit, avec stimulus conditionnel, l'agent incapable de pseudo-conditionnement devrait monter en permanence, et cela de manière identique entre les deux configurations, car il ne répondrait pas au stimulus. Par contre, un agent capable de pseudo-conditionnement devrait commencer par descendre comme s'il s'agissait d'une stimulation avec stimulus inconditionnel, puis progressivement se remettre à monter lorsque l'effet du pseudo-conditionnement s'estompe.

Ces graphiques ne sont pas les résultats de simulation mais sont "dessinés à la main" pour illustrer les propriétés du pseudo-conditionnement.

### 4.3.1 Modèle avec masque sur les nœuds

Nous ajoutons un masque  $\mathbf{N}_m$  de dimension  $K$  (nombre de dimensions de l'espace sensorimoteur) avec des composantes dans  $[0, 1]$  comme nouvelle composante à chaque nœud  $N$ .  $\mathbf{N}_m$  est utilisé pour calculer une distance virtuelle *masquée*  $d_m$  entre le nœud  $N$  et la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$  :

$$d_m(\mathbf{N}_p, \mathbf{x}_{SM}) = \frac{2}{1 + \exp(k_d \|\mathbf{N}_p - \mathbf{x}_{SM}\|_m^2)} \quad (4.27)$$

$$\|\mathbf{N}_p - \mathbf{x}_{SM}\|_m^2 = \sum_i \mathbf{N}_{mi}^2 (\mathbf{N}_{pi} - \mathbf{x}_{SMi})^2 \leq \|\mathbf{N}_p - \mathbf{x}_{SM}\|_2^2 \quad (4.28)$$

La distance  $d_m$ , diminuant artificiellement la distance sensorimotrice entre l'agent et les nœuds,

est utilisée à la place de  $d$  dans les équations 3.7 et 3.8.

Il s'ensuit que pour une dimension  $i$  donnée, la composante de masque associée la plus spécifique est  $\mathbf{N}_{m_i} = 1$  (cas où la distance entre  $\mathbf{N}_{\mathbf{p}_i}$  et  $\mathbf{x}_{SM_i}$  est entièrement prise en compte). Le plus général que puisse être la composante de masque associée est  $\mathbf{N}_{m_i} = 0$  (cas où la distance entre  $\mathbf{N}_{\mathbf{p}_i}$  et  $\mathbf{x}_{SM_i}$  est complètement ignorée, comme si on avait  $\mathbf{N}_{\mathbf{p}_i} = \mathbf{x}_{SM_i}$ ).

À sa création, un nœud a comme masque  $\mathbf{1}$  (vecteur dont toutes les composantes valent 1), le plus spécifique possible. Ensuite, le masque évolue selon :

$$f_{gen} = (\sqrt{K^\sigma} - \|[\mathbf{x}_{SM}]^\sigma - [\mathbf{N}_{\mathbf{p}}]^\sigma\|)\beta_m \cdot \omega(N_w) \cdot d_m(\mathbf{N}_{\mathbf{p}}, \mathbf{x}_{SM}) \quad (4.29)$$

$$\frac{\partial \mathbf{N}_m}{\partial t}(\mathbf{x}_{SM}, t) = \gamma_m \cdot \mathbf{1} - f_{gen} \cdot \mathbf{N}_m \quad (4.30)$$

avec  $K^\sigma$  le nombre de dimensions sensorielles (soit le nombre de senseurs de l'agent). Les paramètres  $\beta_m$  et  $\gamma_m$  régulent respectivement la généralisation et la spécialisation du masque. Ils sont fixés par défaut à  $\beta_m = 3$  et  $\gamma_m = 0.05$ . Chacune des composantes du masque est toujours conservée dans  $[0, 1]$ .

Ainsi, plus la distance masquée entre un nœud  $N$  et la position sensorimotrice de l'agent  $\mathbf{x}_{SM}$  est petite, plus les composantes du masque  $\mathbf{N}_m$  sont diminuées (et donc  $\mathbf{N}_m$  se généralise). La spécialisation (ou augmentation des composantes) du masque se produit légèrement et continuellement au fil du temps.

La figure 4.30b se base sur l'instanciation de l'IDSM utilisée dans les expériences et décrite en détail dans la section suivante. Elle illustre la modification de la dynamique du sous-espace  $m \times s_1$  lorsque le masque du nœud  $N^1$  est généralisé depuis  $\mathbf{N}_m = [1, 1, 1]$  à  $\mathbf{N}_m = [0.5, 0.5, 0.5]$ . La zone d'influence du nœud  $N^1$  (un bloc de la moitié supérieure de l'espace par rapport à l'axe du senseur  $s_1$ ) est augmentée dans l'ensemble de l'espace excepté un quart d'ellipsoïde avec  $N^2$  comme centre (zone d'influence du nœud  $N^2$  réduite en conséquence).

### 4.3.2 Expériences et résultats

$\Delta t$	$k_t$	$k_d$	$k_\omega$	$\tau$	$\mu$
0.1	1	30	0.025	10	$\in [0; 1]$

TABLE 4.4 – Valeurs des paramètres et options utilisées pour les résultats sur le pseudo-conditionnement (section 4.3.2)

#### Instanciation du modèle

Pour les expériences sur ce modèle, l'agent a un moteur et deux senseurs. Le senseur  $s_1$  perçoit le stimulus inconditionnel  $SI$  et le senseur  $s_2$  perçoit le stimulus conditionnel  $SC$ . Les nœuds de l'espace sensorimoteur  $SM$  sont illustrés sur la figure 4.30a, sans représentation de la dimension du second senseur  $s_2$  pour des raisons de simplicité (nous ne représentons que le sous-espace  $m \times s_1$  en fixant  $s_2 = 0$ ). Les deux nœuds sont définis par :

- $N^1 = \langle \mathbf{p} = [0.5, 1, 0], \mathbf{v} = [-0.5, 0, 0], w = 0, m = \mathbf{1} \rangle$ ,
- $N^2 = \langle \mathbf{p} = [0, 0, 0], \mathbf{v} = [0.5, 0, 0], w = 0, m = \mathbf{1} \rangle$ .

Notre objectif avec ces deux nœuds est de produire un comportement par défaut consistant à se déplacer vers le haut sans stimulation et vers le bas lorsque le stimulus inconditionnel  $SI$

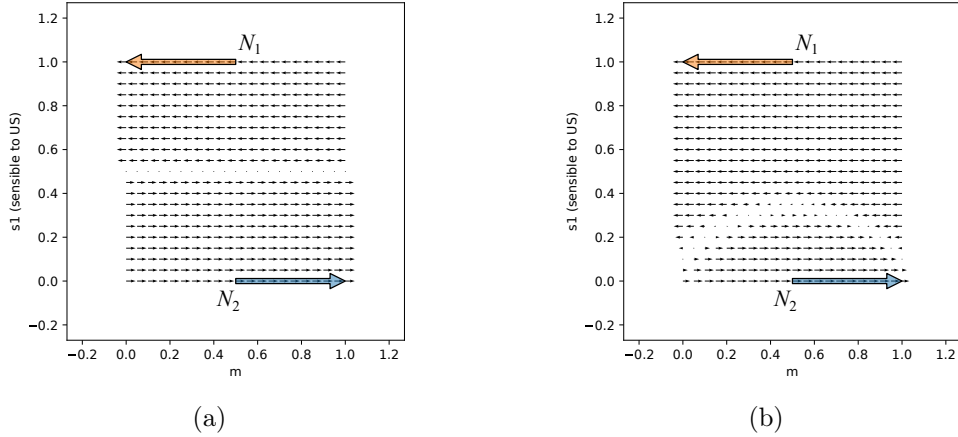


FIGURE 4.30 – Variation de la commande motrice ( $d\mu$ ) dans le sous-espace sensorimoteur  $m \times s_1$  d'un IDSMS avec deux nœuds utilisés dans les expériences sur le pseudo-conditionnement. Les grandes flèches sont les nœuds, la base étant le  $N_p$  et la longueur de la flèche représentant le  $N_v$ . a) : le masque de  $N^1$  est  $N_m = [1, 1, 1]$ . b) : avec un masque  $N_m = [0.5, 0.5, 0.5]$ , cela conduit au renforcement de son influence face à  $N^2$ .

(perçu par le capteur  $s_1$ ) est activé. Le stimulus conditionnel  $SC$  étant neutre, la dynamique est inchangée sur la dimension du capteur  $s_2$ . La dynamique se résume donc à deux blocs, l'un au-dessus du plan  $s_1 = 0,5$ , avec le comportement descendant activé par le  $SI$ , et l'autre en dessous du plan  $s_1 = 0,5$ , avec le comportement ascendant.

### Configuration expérimentale

Afin de mesurer si nos modèles sont capables de pseudo-conditionnement, nous utilisons le protocole suivant. Ce que nous cherchons à évaluer est la capacité du modèle à produire une réponse généralisée à un stimulus préalablement neutre (le stimulus conditionnel  $SC$ , perçu par le capteur  $s_2$ ), après présentation continue ou répétée d'un stimulus non neutre (le stimulus inconditionnel  $SI$ , perçu par le capteur  $s_1$ ).

L'expérience est divisée en trois phases de vingt unités de temps. Durant la première phase, le  $SC$  est présenté, afin d'observer le comportement de l'agent avant le pseudo-conditionnement. Le  $SC$  est ensuite remplacé par le  $SI$  lors de la deuxième phase. Enfin, le  $SC$  est présenté à nouveau lors de la troisième phase, ce qui devrait provoquer une réponse de l'agent similaire à celle de la seconde phase s'il y a eu pseudo-conditionnement. Lorsqu'un stimulus est actif, la valeur du capteur correspondant est fixée à 1. Pour limiter les différences entre les conditions expérimentales, au début de chaque phase, l'agent est réinitialisé à une position de départ à vitesse nulle ( $x_E = \dot{x}_E = \ddot{x}_E = 0$ ) et à un état sensorimoteur  $x_{SM} = [0.5, 0, 0]$ . Ce choix de forcer la valeur du capteur à 0 est fait pour avoir des conditions initiales identiques entre les phases, en particulier pour comparer la première et la troisième phase. Comme expliqué dans la section 4.3.1, l'IDSMS est modifié en conséquence.

Dans une expérience donnée, la stimulation suit l'un des quatre schémas ci-dessous. Pour une expérience donnée, toutes les stimulations (qu'elles soient  $SC$  ou  $SI$ ) suivent le même schéma.

- (a) : stimulus continu ;
- (b) : répétitions de 3 unités de temps avec stimulus / 1 unité de temps sans ;

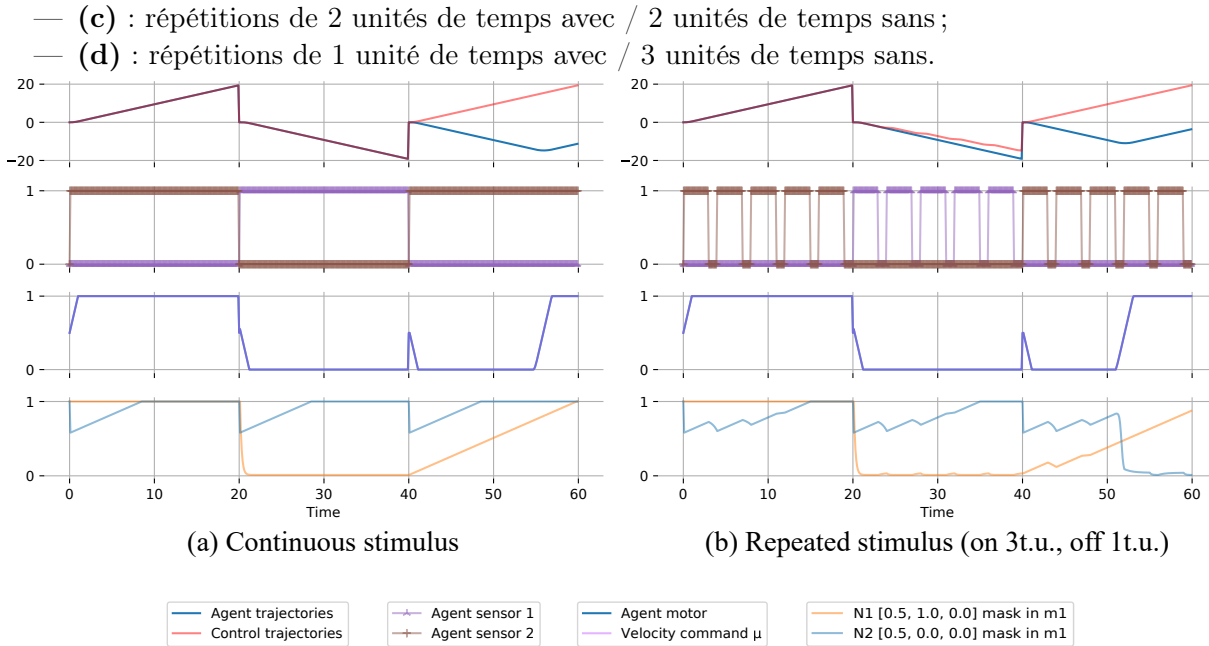


FIGURE 4.31 – Résultats pour les schémas (a) et (b) pour tester s’il y a pseudo-conditionnement. De haut en bas : trajectoires des agents apprenant et témoin, valeurs des senseurs, valeurs de la composante motrice  $[\mathbf{x}_{SM}]^\mu$  et valeurs des masques. Lors de la troisième phase (après 40u.t.), la généralisation plus forte de  $N^1$  par rapport à  $N^2$  (voir le graphique des masques) conduit à un comportement pseudo-conditionné de l’agent : il apprend à se déplacer vers le bas lorsque le  $SC$  est perçu.

## Résultats

Les résultats sont illustrés sur les figures 4.31 et 4.32. Pour l’agent témoin, le comportement par défaut (mouvement vers le haut) est observé pendant la première et la troisième phase, lors de la présentation du  $SC$ . Pendant la deuxième phase, avec la présentation du  $SI$ , le comportement de réponse (mouvement vers le bas) est observé.

Il y a une sous-figure pour chacun des schémas de stimulation, avec la même disposition de graphiques. Le graphique du haut affiche la trajectoire de l’agent apprenant et celle de l’agent témoin. Le second graphique en partant du haut donne des informations sur la perception des deux stimuli par l’agent (valeurs des senseurs). Le troisième graphique en partant du haut montre la valeur motrice normalisée de l’agent dans l’espace sensorimoteur (la première composante de  $\mathbf{x}_{SM}$ ), et la commande motrice  $\mu$ . Afin de ne pas surcharger inutilement la figure, le dernier graphique montre l’évolution des masques des nœuds réduits à leur seule composante motrice. En effet, pour un même masque, les autres composantes ont exactement les mêmes valeurs.

Avec tous les schémas de stimulation, le comportement de l’agent d’apprentissage pendant la troisième phase (présentation du  $SC$  suivant la présentation du  $SI$ ) est un mélange entre ses comportements de la première (comportement ascendant) et de la deuxième phase (comportement descendant).

Pour le schéma (a), le pseudo-conditionnement n’a aucune influence sur le comportement observé dans les première et deuxième phases. Dans la dernière phase, le pseudo-conditionnement



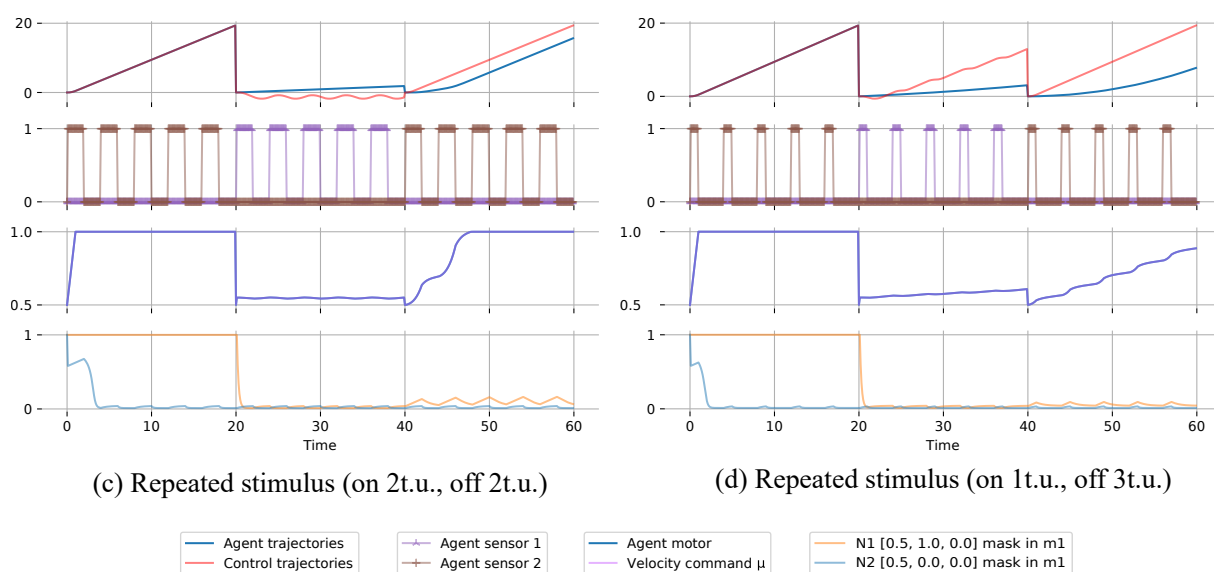


FIGURE 4.32 – Résultats pour les schémas (c) et (d) pour tester s’il y a pseudo-conditionnement. De haut en bas : trajectoires des agents apprenant et témoin, valeurs des senseurs, valeurs de la composante motrice  $[x_{SM}]^\mu$  (attention à l’échelle par rapport aux schémas (a) et (b), ici la valeur minimale est 0.5) et valeurs des masques. Ces schémas mettent en évidence que pendant la phase 2 (entre 20u.t. et 40u.t.), la généralisation des masques conduit à des contributions similaires des deux nœuds et donc à l’immobilité de l’agent. En phase 2 et 3, les valeurs des masques inhibent fortement l’expression du comportement de descente, le moteur de l’agent met seulement plus ou moins de temps à atteindre son maximum.

permet à l'agent d'expérimenter un "nouveau" comportement : l'agent descend en ne percevant que le *SC*, alors qu'il serait monté sans cet apprentissage. L'évolution des masques montre que seul le nœud  $N^1$ , responsable de la réponse au *SI*, atteint une généralisation maximum. À tout moment, le masque de  $N^1$  est plus généralisé (ses composantes sont plus proches de 0) que le masque de  $N^2$ . La spécialisation constante du masque de  $N^1$  permet au comportement par défaut de s'exprimer à nouveau à la fin de la troisième phase. Le nœud  $N^2$ , responsable du comportement par défaut de descente, ne se généralise que légèrement et temporairement suite à la réinitialisation de l'agent au début de chaque phase.

Dans tous les schémas de stimulation avec répétition ((**b**), (**c**) et (**d**)), le masque du nœud  $N^2$  se généralise chaque fois qu'il n'y a pas de stimulus actif. Dans le schéma (**b**), avec un temps de désactivation court et d'activation long du stimulus, lors de la seconde phase, l'agent ne retrouve pas le comportement par défaut (monter) à la désactivation du stimulus. La cause en est que le nœud  $N^1$  est plus généralisé que  $N^2$ , l'agent se comporte alors comme si le stimulus inconditionnel *SI* était activé en continu. Lors de la troisième phase, cela reste le cas pendant les deux premières désactivations (43u.t. et 47u.t.). Ensuite, l'équilibre entre la généralisation des nœuds à la troisième désactivation (51u.t.) permet à  $N^2$  de devenir plus influent que  $N^1$ , d'où le changement de comportement : l'agent se remet à monter. Pour les schémas (**c**) et (**d**), l'augmentation du temps hors stimulation permet la généralisation maximale de  $N^2$  dès la première phase. Une conséquence est la quasi immobilité de l'agent pendant la seconde phase : les deux nœuds ayant une d'influence similaire, la variation de la commande motrice  $\mu$  devient presque nulle. Dans les deux schémas, pour les phases deux et trois, la composante motrice  $[\mathbf{x}_{SM}]^\mu$  ne descend jamais en dessous de 0.5 et par conséquent le comportement descente n'est jamais exprimé. La différence dans la troisième phase, où l'agent monte plus rapidement dans le schéma (**c**) que dans le (**d**) provient du fait que dans cette phase,  $N^1$  est généralisé dans les périodes hors stimulation et spécialisé lorsque le stimulus conditionnel *SC* est présenté. Les périodes hors stimulation étant plus longues et les périodes avec stimulation plus courtes dans le schéma (**d**),  $N^1$  reste plus généralisé qu'en (**c**).

Pour résumer, on observe que dans tous les schémas le comportement résultant en troisième phase est le résultat du rapport d'influence entre la généralisation de  $N^1$  (lorsque *SI* est perçu, ou hors stimulation s'il est déjà assez généralisé) et la généralisation de  $N^2$  (lorsqu'il n'y a pas de stimulation). Nous avons donc un mécanisme non discriminant de généralisation de comportement qui est activé plus ou moins fortement selon la durée d'exposition à un certain contexte sensorimoteur.

### 4.3.3 Bilan sur le pseudo-conditionnement

Nous avons proposé ce modèle pour permettre à un agent basé sur l'IDSMS d'apprendre de manière similaire à du pseudo-conditionnement à travers des masques rajoutés aux nœuds.

La généralisation de l'influence des nœuds à travers les masques entraîne un transfert de la réponse comportementale du stimulus déclencheur (ou stimulus inconditionnel) vers d'autres stimuli. De plus, notre modèle permet également la généralisation de masques de nœuds qui ne sont "liés" à aucun stimuli (les composantes sensorielles de la position de tels nœuds sont toutes fixées à 0). Cependant, nous ne savons pas si cela a un sens du point de vue de la biologie, d'une part par la définition du pseudo-conditionnement qui intervient après qu'un stimulus ait induit un comportement de réponse et d'autre part car la notion d'un état de l'environnement où un organisme n'en percevrait absolument rien semble très étonnante.

Le mécanisme de masque que nous proposons pour le pseudo-conditionnement conduit à une généralisation large : une fois qu'un nœud est entièrement généralisé, il a un impact sur le comportement de l'agent de manière égale partout dans l'espace sensorimoteur. Ce comportement généralisé n'est contrebalancé que par l'existence d'autres nœuds (par le biais des équations de l'IDSM original). Cette "compétition" entre les nœuds peut ne pas être suffisamment forte pour préserver les autres comportements existants. Dans un tel cas, il peut être nécessaire d'introduire un mécanisme de spécialisation façonnant les régions sensorimotrices où le nœud doit effectivement se généraliser, ou de restreindre les dimensions sur lesquelles il peut le faire. Par exemple, la spécialisation pourrait être déclenchée par la divergence entre la composante  $\mathbf{N}_v$  du nœud  $N$  et le dernier mouvement de l'agent dans l'espace sensorimoteur  $SM$ , puisqu'une telle différence signifierait que la commande motrice stockée dans  $N$  n'est pas en adéquation avec l'environnement actuel. L'ajout d'un autre apprentissage non-associatif comme l'habituation pourrait également contribuer à prévenir une "surgénéralisation" en minimisant l'impact des répétitions de stimuli de faible intensité.

Alternativement à ce modèle des masques, toujours dans le contexte de l'IDSM, nous aurions pu proposer d'autres mécanismes permettant un pseudo-conditionnement proche de celui expérimenté ici. L'objectif étant d'adapter la région d'influence sensorimotrice d'un nœud, nous aurions pu définir et adapter un  $k_d$  par nœud, ce qui aurait donné des résultats similaires. Cependant, le mécanisme du masque est strictement plus général dans le sens où il permet (bien que non illustré ici) de généraliser selon certaines dimensions tout en restant spécifique selon d'autres. Par ailleurs, le choix de n'appliquer la distance masquée  $d_m$  uniquement pour la densité des nœuds actifs  $\phi_A$  (équation 3.7) et pour le vecteur de vitesse  $\mathbf{V}$  (équation 3.8), donc ne prenant pas en compte le vecteur d'attraction  $\mathbf{A}$  (équation 3.9), peut être discuté. En effet, ce choix repose sur l'idée de ne généraliser que la composante de vitesse d'un nœud et non pas son pouvoir attractif. Dans le cas où la création de nœuds est activée, la distance masquée peut aussi être appliquée à la densité  $\phi_N$  d'une part pour que la création de nœuds dans une zone peu dense dépende de la généralisation des anciens nœuds, mais aussi pour réguler le nombre de nœuds à créer dans une même zone.

Les résultats de nos expériences dépendent fortement de nombreux paramètres. Le nombre de nœuds, leurs positions et les valeurs des différents paramètres IDSM sont très interdépendants et ont été soigneusement choisis. Nous avons considéré un IDSM très simple pour expérimenter le mécanisme du masque et des configurations plus complexes avec plus de nœuds et des comportements résultants eux-aussi plus complexes restent à étudier. De plus, le modèle original IDSM inclut une évolution des poids et la génération de nouveaux nœuds qui ont été ici supprimés pour faciliter l'analyse. L'idée à terme est que le mécanisme de masque doit permettre à l'agent d'expérimenter des comportements connus dans des situations inconnues, d'influencer les comportements normalement par défaut de ces zones par ceux qui ont été récemment expérimentés par l'agent, puis finalement de mémoriser la dynamique résultante par la création de nouveaux nœuds.

## 4.4 Bilan de ces modèles

En bilan de ce chapitre, nous faisons le point sur chacun des quatre modèles présentés : modèle avec mémoire sensorielle, modèle avec impact sur l'énergie des nœuds, modèle avec impact sur le poids des nœuds et modèle avec masques. En particulier, après une synthèse de leurs caractéristiques, nous les replaçons dans la problématique de la thèse : est-ce que ces modèles permettent

à un agent d'étendre son répertoire comportemental ?

**Modèle avec mémoire sensorielle** Ce modèle permet de mettre en œuvre habituation et sensibilisation spécifique au site quand le stimulus est présenté de manière continue.

En fait, si l'on suppose qu'une réponse à un stimulus est dépendante de l'intensité du stimulus, que la réponse est plus forte pour un stimulus plus intense, alors le modèle avec mémoire sensorielle adapté pour l'habituation ou la sensibilisation spécifique au site permet à l'apprentissage en question d'avoir lieu. Cependant, cette dépendance n'est pas garantie. Avec une mémoire sensorielle, l'agent montre de l'habituation lorsque sa position sensorimotrice s'éloigne de la dynamique responsable de la réponse et/ou se rapproche d'une dynamique différente, et inversement pour la sensibilisation spécifique au site. Cela ne dépend pas de l'intensité du stimulus mais uniquement de la disposition de la dynamique. En particulier, on pourrait imaginer une configuration où l'inhibition du senseur amène à un renforcement de la réponse et inversement. Pour résumer l'effet de ce modèle : lorsque le stimulus est perçu, la version habituation force l'agent à aller progressivement vers la zone à valeur sensorielle nulle là où la version sensibilisation le pousse vers la zone à forte valeur sensorielle.

Pour notre problématique, ce que nous retenons est surtout qu'une variation de la position sensorielle dans un espace sensorimoteur avec dynamique variée amène à une modification de comportement. Dans l'exemple simple avec deux moteurs et un senseur permettant de séparer deux comportements (section 3.3), un passage progressif de  $s = 1$  à  $s = 0$  permettrait d'exprimer jusqu'à trois comportements (le premier, un mélange des deux, le second). De plus, à l'inverse de nos autres modèles, la mémoire sensorielle n'impacte pas directement la dynamique de l'espace sensorimoteur, ce qui simplifie l'interprétation des résultats lorsqu'elle est utilisée surtout avec de nombreux nœuds. Cependant, elle a quand même un impact indirecte lorsque l'IDSM déformable est utilisé : à travers la création des nœuds sur la trajectoire sensorimotrice modifiée par la mémoire sensorielle et à travers la modification des poids (affaiblissement des nœuds dont on force l'éloignement, renforcement des nœuds dont on force le rapprochement).

Un autre point important est que les configurations dans lesquelles le modèle avec mémoire sensorielle a été testé sont très simples, le stimulus étant binaire (soit actif, soit inactif). Il serait intéressant d'observer l'évolution de la mémoire sensorielle et le comportement résultant en utilisant une source stimulante fixée dans l'environnement et qui générerait ainsi une valeur sensorielle brute perçue dont la valeur dépendrait de la position de l'agent par rapport à cette source. En particulier, il serait d'observer ce qui se passe quand la mémoire sensorielle est saturée (l'agent est proche du stimulus pendant un certain temps) et que l'intensité du stimulus baisse subitement sans toutefois devenir nulle. Est-ce que, par exemple, l'agent a eu le temps de s'habituer (au sens de l'IDSM) à cette haute intensité ? Et dans ce cas, peut-il alors être touché par de l'habituation à basse intensité ? Dans le modèle actuel, cela dépendrait des valeurs de haute et basse intensité du stimulus, et des paramètres du modèle (paramètres fixés "arbitrairement").

**Mécanisme de surprise** Ce mécanisme, qu'il impacte l'énergie de nœuds ou leurs poids, permet de mettre en œuvre habituation et sensibilisation spécifique au site quand le stimulus est présenté de manière répétée, car il est sensible aux *variations* du stimulus. En effet, la surprise dépend du *changement* d'intensité plutôt que de l'intensité elle-même.

Ce mécanisme en lui-même nous semble intéressant car il permet une modification du com-

portement "par défaut", c'est-à-dire du comportement qui est exprimé quand il n'y a pas de stimulation. En effet, lorsqu'il y a de la surprise, *tous* les nœuds sensorimoteurs proches de l'actuelle position sensorimotrice de l'agent sont impactés, qu'ils participent au comportement réponse ou au comportement par défaut.

D'ailleurs une question ouverte subsiste, celle de savoir s'il est plus intéressant de tenir compte de toutes les dimensions sensorimotrices pour déterminer les nœuds impactés (ce qui est fait actuellement) ou s'il serait plus intéressant de ne s'intéresser qu'aux dimensions sensorielles. Par exemple, dans le cas d'une variation positive du stimulus (il passerait de 0 à 1), cette deuxième possibilité permettrait à la surprise d'influencer tous les nœuds liés à la présence de ce stimulus (les nœuds dont la composante sensorielle est proche de 1), quel que soit le comportement moteur actuel de l'agent. Pour le moment il nous est difficile de jugé laquelle des deux possibilités est la plus intéressante, mais elles sont aussi simple l'une que l'autre à mettre en œuvre.

Autre point, dans nos expériences avec la surprise nous n'utilisons jamais plus d'une seule dimension sensorielle. Cependant, l'impact sur un nœud  $N$  n'est pas dépendant d'une variation d'un stimulus spécifique, cet impact ne dépend que d'une différence entre une position sensorimotrice "attendue" et la position actuelle. Par conséquent, une surprise suite à une désactivation d'un stimulus (par exemple) pourrait impacter des nœuds liés à d'autres stimulus si ces derniers sont toujours présents dans l'environnement. Les conséquences de cette caractéristique ne sont pas évidentes à prévoir sans analyses plus poussées, et seront sans doutes très contextuelles.

**Modèle avec impact sur l'énergie** Le modèle avec impact sur l'énergie permet à l'agent d'exprimer des effets d'habituation et de sensibilisation spécifique au site dans nos expériences. Cependant, ces effets découlent du mécanisme de surprise et sont totalement dépendants de la configuration de l'espace sensorimoteur et des paramètres du modèle. En effet, si l'activation (resp. la désactivation) du stimulus ne conduit l'agent assez proche d'un nœud, il n'y aura pas ou peu d'apprentissage à moins de modifier fortement le paramètre  $\beta_e$ . Par exemple, dans les expériences pour la sensibilisation spécifique au site, le stimulus activé à 0.75 résulte en une distance entre l'agent et le nœud  $N^1$  responsable du comportement réponse plus importante que dans les expériences pour l'habituation. Ainsi, il faut un paramètre  $\beta_e$  quatre fois supérieur à celui de l'habituation pour avoir des impacts similaires sur la consommation d'énergie.

En ce qui concerne notre problématique, le mécanisme d'énergie ajouté aux nœuds et le mécanisme de surprise permettent de rajouter un levier sur la dynamique en cas de changement important et brusque de la valeur sensorielle. Cet impact augmente ou réduit l'influence du nœud concerné sans changer la zone d'influence de ce nœud (l'influence des autres nœuds empiète peu dessus). Dans le cas d'une réduction d'influence, le résultat semble similaire à une réduction de la composante  $\mathbf{N}_v$  du nœud, et la zone d'influence autour du nœud peut aller jusqu'à redevenir neutre (comme s'il n'y avait pas de nœud). Selon l'énergie restante du nœud, lorsque l'agent traverse cette zone sensorimotrice, sa commande motrice  $\boldsymbol{\mu}$  en est peu ou pas modifiée. Dans le cas d'un renfort d'influence, tant que l'énergie n'est pas très petite ou nulle, l'effet inverse se produit avec un effet similaire à une forte augmentation de la composante  $\mathbf{N}_v$ , et une zone d'influence inchangée. Par contre, si l'énergie devient trop petite (voire quasiment nulle), nous observons un gain d'influence tel qu'il change aussi la dynamique dans la zone d'influence des autres nœuds. Ce cas très particulier semble dû à des problèmes de précisions dans les calculs, où plusieurs nombres à la limite de ce qui est représentable en virgule flottante sont additionnés. Ce cas particulier expliquerait le phénomène d'oscillations observées lors des expériences avec le mo-

dèle utilisant l'énergie. L'utilisation d'une valeur *minimale* pour l'énergie permettrait peut-être de ne pas avoir ce cas de "débordement".

**Modèle avec impact sur le poids** Comme pour le modèle avec énergie, le modèle avec impact sur le poids utilise le mécanisme de surprise et son effet dépend de la configuration de l'espace sensorimoteur au sens où la position des différents nœuds est importante. Le modèle fonctionne bien pour l'habituation, mais est plus compliqué à calibrer pour la sensibilisation spécifique au site. Cela provient d'une part, comme pour le modèle avec énergie, de la distance minimale plus importante entre l'agent et les nœuds, et d'autre part du fait que les gains et les pertes d'influence ne sont pas symétriques autour d'une valeur de poids de 0 (pour rappel, les poids peuvent être négatifs).

Par rapport à notre problématique d'avoir des mécanismes permettant d'observer des comportements nouveaux ou différents dans un même contexte environnemental, la caractéristique de l'impact sur le poids à pouvoir moduler les *taille* des zones d'influences des nœuds est intéressante. Encore une fois, nos expériences ont été réalisées dans des environnements très simples. Avec des espaces sensorimoteur plus complexes, l'inhibition d'un nœud et de sa zone d'influence conduisant à l'extension d'autres zones d'influences, nous devrions pouvoir observer des cas où des nœuds sans "interactions" initialement en obtiennent. Plus de travaux sont nécessaires pour voir si c'est effectivement le cas. En particulier, il faudrait analyser si l'effet d'une pénalité sur les poids est aussi efficace pour changer le paysage de la dynamique de l'espace sensorimoteur lorsqu'il y a redondance dans l'information de nombreux nœuds.

**Modèle avec les masques** Ce modèle permet de mettre œuvre du pseudo-conditionnement au sein de l'IDSM, en généralisant temporairement le comportement d'un nœud à tout l'espace sensorimoteur.

En effet, le dernier modèle que nous avons présenté, avec le mécanisme de masque, permet de généraliser un nœud  $N$  en diminuant virtuellement la distance entre la position sensorimotrice de l'agent et la position du nœud masqué. Cette généralisation dépend de la distance masquée entre l'agent et le nœud, ainsi que de la proximité entre la position sensorielle de l'agent et de la position sensorielle du nœud (ce qui permet à un nœud généralisé au maximum de pouvoir se re-spécialiser dans un contexte sensoriel différent du sien). Dans la configuration de notre expérience, ce modèle permet à l'agent d'apprendre par pseudo-conditionnement en ré-itérant sa réponse à un stimulus, ce qui devient sa réponse à un autre stimulus.

Le mécanisme de masque est efficace pour étendre significativement la zone d'influence d'un nœud, ce qui est une alternative au mécanisme de poids qui lui est surtout efficace pour réduire cette zone d'influence. Cette modification de la dynamique de l'espace sensorimoteur permet de provoquer des changements de comportements chez l'agent, caractéristique qui nous intéresse. La généralisation d'un nœud pourrait cependant être trop importante, trop "efficace", du fait du masque le plus général rendant virtuellement la position sensorimotrice du nœud masqué identique à celle de l'agent. Dès lors, quelle que soit la position sensorimotrice de l'agent, son comportement est uniquement dû à l'influence de ce nœud au masque le plus généralisé. Le mécanisme tel qu'il est actuellement permet un impact progressif sur l'espace sensorimoteur, comme illustré sur la figure 4.30. À l'inverse, du point de vue d'une position sensorimotrice donnée, de part les caractéristiques du facteur  $d$  (voir équation 3.4 et figure 3.3), l'impact peut être négligeable pour une généralisation du masque déjà bien avancée et brusquement de plus

en plus important en approchant de la généralisation maximale. L'articulation entre le modèle avec masques et l'IDSM déformable (c'est-à-dire un IDSM on l'on peut créer des nœuds et les adapter) n'a pas été testé. L'utilisation d' $\omega(N_w)$  dans la mise à jour du masque laisse le rôle habituel du poids du nœud, permettant un nœud avec poids plus fort d'être plus facilement généralisé et inversement. Dans le cas de l'exploration d'une zone peu dense en nœuds, la création de nœuds permettrait l'enregistrement d'un comportement résultant de la dynamique initiale impactée par la généralisation des nœuds concernés. Si au contraire la zone est dense en nœuds, par exemple en suivant une trajectoire sensorimotrice exprimée auparavant, il est compliqué de prédire le résultat de ce pseudo-conditionnement. Il se peut que la généralisation poussée de certains nœuds éloigne l'agent de cette trajectoire et permette alors la création de nœuds avec un  $N_v$  très différent de celui des nœuds existant auparavant dans cette zone, ce qui conduirait de nouveau à l'enregistrement d'un nouveau comportement.





## Bilan et Perspectives

À travers cette thèse, nous nous sommes intéressés à l'autonomie d'un agent artificiel à travers une question précise : comment un agent peut-il apprendre des comportements sensorimoteurs plus complexes que ceux qu'il possède initialement à partir de ses interactions avec son environnement ?

Après une étude de l'état de l'art sur les notions de comportements et d'apprentissages, nous avons fait le choix de chercher à répondre à cette question à l'aide des apprentissages non-associatifs. Ces apprentissages sont à la fois les plus simples connus et les plus observés dans le vivant, au point que l'habituation a été observée dans des organismes aussi divers que les bactéries (même si la terminologie entre apprentissage ou adaptation est toujours discutée), les ciliés (organismes unicellulaires), les plantes et les animaux.

En nous basant sur le modèle IDSM, nous avons proposé quatre modèles dont les mécanismes permettent à un agent simulé d'exprimer les différents apprentissages associatifs. Trois modèles permettent l'habituation et peuvent être modifiés pour permettre à la place la sensibilisation spécifique au site. Le quatrième modèle a été conçu pour le pseudo-conditionnement.

Dans chaque cas, ces apprentissages permettent effectivement de modifier le comportement "réflexe" de l'agent pour répondre différemment à un certain contexte environnemental, selon les caractéristiques de l'apprentissage considéré.

Les différents modèles proposés sont testés, de manière expérimentale, dans des configurations simples permettant l'observation des effets des différents apprentissages non-associatifs. Chaque modèle impacte différemment le calcul de la variation de la commande motrice et modifie ainsi plus ou moins directement le comportement résultant de l'agent. Le modèle avec mémoire sensorielle modifie la valeur du stimulus perçu par le senseur (ou valeur sensorielle brute  $s_b$ ), modifiant la position sensorielle de l'agent dans l'espace sensorimoteur. Le modèle avec impact sur l'énergie des nœuds modifie l'énergie, composante ajoutée à chaque nœud et utilisée dans les vecteurs vitesse  $\mathbf{V}$  et attraction  $\mathbf{A}$ , mais pas dans les densités  $\phi_{\mathcal{N}}$  et  $\phi_{\mathcal{A}}$ , ce qui résulte en un renfort ou affaiblissement de l'influence des nœuds mais sans modification significative sur leur zone d'influence. Le modèle avec impact sur le poids des nœuds utilise la composante de poids  $N_w$  déjà existante dans les nœuds de l'IDSM, utilisée dans les vecteurs  $\mathbf{V}$  et  $\mathbf{A}$ , et dans les densités  $\phi_{\mathcal{N}}$  et  $\phi_{\mathcal{A}}$ . Le résultat est une modification de la zone d'influence des nœuds, particulièrement

propice à la réduction de la zone du nœud avec poids réduit et à l'augmentation de celles des nœuds non impacté. Enfin, le modèle avec masques permet une généralisation d'un nœud visité régulièrement à travers une distance masquée  $d_m$  utilisée dans le vecteur vitesse  $\mathbf{V}$  et dans la densité  $\phi_{\mathcal{A}}$ , réduisant virtuellement la distance entre la position sensorimotrice de l'agent et le nœud généralisé et augmentant ainsi la zone d'influence du nœud.

Nous avons montré comment ces différents modèles pouvaient doter l'IDSM de mécanismes d'apprentissage non-associatif. En particulier, la mémoire sensorielle semble adaptée pour apporter une habitude ou une sensibilisation spécifique au site dans le cas de stimuli continus. Les modèles modifiant les composantes d'énergie ou de poids des nœuds sont plus pertinents pour observer de l'habitude et de la sensibilisation spécifique au site dans le cas de stimuli répétés. Enfin, le modèle qui ajoute un masque aux nœuds permet de mettre en œuvre du pseudo-conditionnement. Nous avons aussi analysé et discuté ces modèles sous le prisme du potentiel qu'ils avaient à doter un agent artificiel de la capacité à étendre son répertoire comportemental (voir section 4.4).

Nous avons par ailleurs bien conscience que certains de nos résultats sont encore préliminaires, et que de nombreuses autres expériences sont nécessaires pour mieux évaluer la portée de nos contributions. Cependant, nos modèles permettent déjà de dégager des pistes de réflexions et des perspectives intéressantes pour des travaux ultérieurs. Nous présentons ces perspectives en commençant par celles que nous estimons relever de considérations à plus court terme, pour terminer par celles relevant plus du long terme.

**Simulation dans des conditions plus complexes** Travailler avec des configurations plus complexes permettrait d'avancer selon deux directions. D'une part, cela serait l'occasion de comprendre les opportunités, mais aussi les problèmes, offerts par le fait de disposer de plusieurs dimensions sensorielles. D'autres part, cela permettrait d'explorer comment se combinent nos apprentissages et comment ils tirent profit, ou non, d'une bibliothèque de comportements "initiaux" plus étoffée.

À court terme, les modèles doivent être testés en augmentant le nombre de dimensions motrices et/ou en utilisant des stimuli avec des sources localisées dans l'environnement permettant une variation continue de la valeur sensorielle. Ces tests devraient être effectués avec l'IDSM simplifié dans un premier temps, puis en ajoutant la création de nœuds et la modification de poids dans un deuxième temps. Par exemple, de nouvelles expériences pourraient se baser sur un agent similaire à celui de la section 3.2.2 avec deux moteurs et deux senseurs. Une question serait alors comment articuler la mémoire sensorielle avec ces différents senseurs : vaut-il mieux une mémoire sensorielle par senseur ou plutôt une mémoire sensorielle "par stimulus", commune à tous les senseurs captant une même source de stimulation ?

En parallèle de la complexification de la configuration, les comportements "initiaux" (qu'ils soient prédéfinis via des nœuds manuellement disposés ou préalablement acquis par l'agent via une marche forcée) doivent eux aussi être complexifiés. Nos modèles pourraient être testés sur des comportements "réflexes" de type taxie (aller vers une source de stimulation) ou phobie (s'éloigner ou éviter une telle source), avec des senseurs différents pour gérer des stimuli différents (une lumière rouge et une lumière verte par exemple). À partir de là nous pourrions traiter des questions intéressantes qui n'ont pas été abordées dans nos travaux. Par exemple, comment s'articulent les apprentissages non-associatifs habitude et sensibilisation spécifique au site pour ne pas s'annuler mutuellement pour un même stimulus ? Est-ce que l'agent pourrait apprendre

---

à s'habituer à la source de phobie et à se sensibiliser à la source de taxie sans guidage explicite, seulement avec une distinction sur l'intensité du stimulus? De plus, d'autres comportements "réflexes" existent dans le vivant, comme par exemple la kinésie : modification de la vitesse ou de l'orientation de déplacement en réponse à la perception du stimulus, mais sans dépendance à la localisation de la source du stimulus (BENHAMOU et BOVET, 1989). Nous n'avons pas encore exploré l'état de l'art à ce sujet, mais il semblerait que ces comportements permettent indirectement à un organisme de passer plus de temps dans sa zone de confort, élément qui pourrait être intéressant sur le long terme dans le cadre de l'autonomie d'un agent artificiel.

Par ailleurs, une réflexion doit aussi être menée en amont ou en parallèle sur l'acquisition de ces comportements "initiaux", réponses aux stimuli pouvant être modifiées par apprentissages non-associatifs. Dans nos expériences, nous avons soit repris la marche forcée pour que l'agent ait une première expérience de son environnement et pour que les nœuds correspondant soient créés dans l'espace sensorimoteur (comportements combinés, voir section 3.3), soit disposé directement les nœuds que nous souhaitons lors de la création de l'agent (toutes les expériences avec nos modèles proposés, voir section 4). Cette dernière méthode serait plus laborieuse et difficilement généralisable dans le cas de comportements plus complexes dans un espace sensorimoteur de plus grande dimension. Une autre possibilité pourrait être de s'appuyer sur les recherches sur les comportements d'exploration. Par exemple, les articles de PYKE, 2015 et WOSNIACK et al., 2017 discutent l'utilisation de la marche (ou du vol) de Lévy ("*Lévy walk*" ou "*Lévy flight*") chez les organismes lorsqu'ils recherchent de la nourriture ("*foraging*") ou autres ressources d'intérêts. Cette marche est un déplacement aléatoire permettant de parcourir de grandes distances, cela pourrait être un comportement "initial" intéressant selon les objectifs donnés à l'agent. Des travaux récents sur l'IDSM utilisent aussi de l'aléatoire pour établir la variation de commande motrice, proportionnellement à une faible densité des nœuds (voir M. EGBERT, 2018 et ZARCO et M. EGBERT, 2019).

**Quelles modifications possibles pour les modèles proposés?** Le modèle avec masques des nœuds a été conçu pour doter l'agent de pseudo-conditionnement. C'est pourquoi nous ne l'avons ni testé pour l'habituation ni pour la sensibilisation spécifique au site. Cependant cela devrait être possible avec peu de modifications : les masques permettent de modifier les zones d'influences des nœuds et notre modèle avec impact sur les poids des nœuds permet ces apprentissages à travers, lui aussi, une modification de la zone d'influence des nœuds. Réfléchir à adapter les masques pour l'habituation et la sensibilisation spécifique au site pourrait être utile face à des variations continues de valeur sensorielle (par exemple cas où l'agent se déplace et que la valeur sensorielle perçue dépend de la distance entre l'agent et la source d'un stimulus continu). En effet, contrairement aux modèles utilisant le mécanisme de surprise qui eux sont adaptés aux variations brusques d'intensité de stimuli (un bruit soudain ou l'activation d'une lampe par exemple), le modèle avec masques permettrait une modification sur les nœuds basée uniquement sur la proximité entre la position sensorimotrice de l'agent et la position des nœuds. Concernant la sensibilisation spécifique au site, nous devrions déjà pouvoir observer une sensibilisation similaire entre les deux modèles (avec masques et avec impact sur le poids des nœuds) dans les conditions de notre expérience sur cet apprentissage. Une configuration plus propice à l'extension de la zone d'influence, point fort de ces modèles, pourrait aussi être testée à la place. Il reste aussi à étudier comment la position relatives des nœuds modifie la forme et la position de la zone d'équilibre de ces influences, car nous avons vu l'importance de cette frontière dans les comportements résultant des apprentissages. Par contre, pour l'habituation, il faudrait modifier le modèle avec masques pour permettre l'augmentation virtuelle de la distance entre la position

sensorimotrice de l'agent et le nœud. Ainsi, nous devrions observer une réduction de la zone d'influence du nœud impacté, ce qui résulterait de l'habituation dans les conditions de notre expérience.

Concernant le modèle avec impact sur l'énergie des nœuds, les expériences sur la sensibilisation spécifique au site mettent en lumière un cas particulier : lorsque l'énergie des nœuds devient nulle, leur influence respective permet de modifier l'état moteur d'une limite à l'autre instantanément. Ce cas nous semble problématique étant donné l'instabilité que cela produit dans le comportement de l'agent et surtout, sur plus long terme avec une instanciation sur un vrai robot, qu'un moteur physique ne pourrait pas faire des écarts aussi importants. Dans le cas de l'habituation, une énergie nulle entraîne une annulation de l'influence du nœud, ce qui ne semble pas problématique car dans la dynamique cela se traduit par une zone "neutre" où la trajectoire motrice de l'agent conserve sa vitesse. Dans tous les cas, si nécessaire, il suffirait d'ajouter un paramètre pour l'énergie minimale pour rendre ces situations extrêmes non atteignables tout en conservant les effets escomptés de l'inhibition et du renforcement de l'influence d'un nœud à travers l'énergie.

**Combinaisons des apprentissages non-associatifs** Ces travaux de thèse ont étudié les différents apprentissages non-associatifs de façon individuelle, pour évaluer les effets des différents modèles concernés. Nous avons parfois combiné certains modèles entre eux (mémoire sensorielle et impact sur l'énergie des nœuds ou mémoire sensorielle et impact sur les poids des nœuds) mais là aussi toujours dans le cadre d'un seul apprentissage, pour tester si ces combinaisons permettaient de montrer l'apprentissage concerné face à nos différents schémas de stimulation.

Dans l'optique de doter un agent de plusieurs de nos modèles avec l'objectif qu'il soit ainsi capable de plusieurs apprentissages non-associatifs, nous sommes confiants sur les possibilités de combinaisons des modèles sans modifications dans une certaine mesure. En effet, chaque modèle est basé sur des composantes différentes, donc utiliser par exemple une mémoire sensorielle pour l'habituation, un impact sur l'énergie pour la sensibilisation spécifique au site et les masques pour le pseudo-conditionnement ne provoque pas de conflits **directs**. Cependant, comme nous avons déjà pu le voir lors de nos expériences avec combinaisons, la mémoire sensorielle impacte la valeur sensorielle prise en compte par l'IDSM, valeur à son tour utilisée par la surprise impactant l'énergie. Il est entendu que ce genre de synergie peut aussi amener à devoir effectuer des modifications sur les modèles pour permettre leur combinaison.

De plus, l'utilisation simultanée de modèles pour l'habituation et de modèles pour la sensibilisation spécifique au site peut être compliquée à paramétrer du fait que les effets de ces deux apprentissages sont l'inverse l'un de l'autre. En biologie, il semblerait que l'habituation soit plus significative pour les stimuli de faible intensité, et la sensibilisation spécifique au site pour les stimuli de forte intensité.

Pour la mémoire sensorielle, on pourrait imaginer un seuil en dessous duquel le stimulus tend à être ignoré (habituation) et au dessus duquel il tend à être exacerbé (sensibilisation spécifique au site). À quel point un tel système pourrait servir au niveau applicatif (réduction de bruit, détection de nouveauté, système d'attention, etc.) reste à déterminer.

Une autre piste pour les modèles utilisant le mécanisme de surprise pourrait être de rendre l'habituation et la sensibilisation spécifique au site dépendantes aux *changements* d'intensité (respectivement aux faibles changements et aux forts changements d'intensité) pour avoir une indépendance à la valeur de l'intensité du stimulus. Cela se traduirait par un seuil sur le méca-

---

nisme de surprise, si la surprise est supérieure au seuil le nœud serait renforcé et inversement.

**Réflexions sur les apprentissages non-associatifs** Nous avons encore d'autres questionnements sur les apprentissages non-associatifs, certains qui pourraient déboucher sur des modifications de modèle individuel et d'autres plus larges les mettant en perspectives des apprentissages associatifs.

Par exemple, pour le pseudo-conditionnement, il n'est pas clair pour nous à quel degré la généralisation se fait en biologie. En particulier, il nous semble peu probable que toute réponse soit généralisable à tout stimulus. La question serait alors sur quels critères baser la restriction de notre mécanisme de masque à certain stimuli plutôt qu'à d'autres.

De plus, dans certaines définitions des apprentissages non-associatifs, il est question des notions de stimulus "significatif pour l'organisme" ou encore "nocif pour l'organisme" ("*noxious*", voir par exemple BYRNE et Robert D HAWKINS, 2015). Ces notions nous interrogent aussi : dans quelle mesure, dans le cadre d'un agent artificiel, une telle distinction est innée ou à acquérir ? Ce qui nous importe essentiellement est à quel point le développeur de l'agent donne des éléments à ce dernier sur l'interprétation de son environnement avant que l'agent ne l'explore de lui-même ? Dans quelle mesure cette distinction est nécessaire à l'agent, pour quels comportements ? Pour le moment, nous n'utilisons dans nos modèles aucune notion de *récompense* qui permettrait à l'agent de distinguer un *bon* d'un *mauvais* stimulus qu'il pourrait alors préférer de suivre ou éviter.

Enfin, concernant les applications possibles de nos modèles d'apprentissages non-associatifs pour un agent artificiel, au moins deux pistes semblent intéressantes à explorer.

D'une part, quelle expressivité comportementale découle d'un modèle d'agent doté uniquement de mécanismes d'apprentissages non-associatifs ? En biologie, de multiples organismes semblent vivre leur vie de manière autonome avec comme apprentissages disponibles uniquement ces apprentissages non-associatifs (cas des organismes unicellulaires par exemple, en l'état actuel de ce qui a été observé chez de tels être vivants). Est-ce alors envisageable d'obtenir un niveau d'autonomie similaire pour un robot doté de mécanismes d'apprentissages non-associatifs mais dépourvu d'apprentissages plus complexes ? Cela dépend aussi bien entendu de la formalisation qui serait faite de l'autonomie. Cependant, la théorie de l'homéostasie comportementale (voir par exemple Edward M EISENSTEIN et al., 2012) est une approche qui pourrait donner des éléments de réponses à ces questions. Par définition, l'homéostasie est le phénomène qui permet le maintien des fonctions vitales de l'organisme. Souvent, ce phénomène est basé sur des processus internes de l'organisme. La théorie de l'homéostasie comportementale rajoute que les modifications de comportements en réponse aux changements environnementaux (augmentation de la température entraînant un déplacement dans une zone ombragée par exemple) sont eux aussi des processus critique pour l'homéostasie, complémentaires aux effets des processus internes. Par conséquent, l'habituation et la sensibilisation seraient alors des apprentissages critiques pour cette homéostasie, et donc pour l'organisme, car ce sont l'une des sources de ces modifications comportementales. Ces apprentissages permettent en particulier à l'organisme une adaptation plus rapide à l'environnement changeant, ainsi qu'une économie d'énergie qui sert elle aussi au maintien des fonctions vitales de l'organisme à plus ou moins long terme.

D'autre part, les apprentissages non-associatifs sont des mécanismes qui peuvent être complémentaires aux apprentissages de plus haut niveau. Ainsi, dans quelle mesure le conditionnement

classique ou d'apprentissage par renforcement implémenté dans un agent capable d'apprentissages non-associatifs bénéficierait des mécanismes d'apprentissages non-associatifs? Des éléments de réponses peuvent être trouvés dans les travaux de MCSWEENEY et MURPHY, 2009. Ils émettent l'hypothèse que la sensibilisation augmente, et que l'habituation diminue, l'efficacité des *renforceurs* ("*reinforcers*", stimuli renforçant les réponses qui les précèdent, potentiellement provoqués par ces réponses, des récompenses par exemple) dans le contrôle du comportement lorsque ces renforceurs sont présentés de manière répétée ou pendant une période prolongée. Pour appuyer cette hypothèse, ils présentent des résultats où l'habituation est en partie responsable de l'arrêt de trois comportements motivés : l'alimentation (habituellement attribuée à la satiété), la course (habituellement attribuée à la fatigue) et la prise de drogue (habituellement attribuée à des facteurs pharmacodynamiques). Des mécanismes artificiels d'apprentissages non-associatifs pourraient peut-être jouer un rôle dans le dilemme exploration/exploitation en robotique, à travers une motivation exacerbée pour un stimulus ou une lassitude pour un autre.

**L'IDSM, et après ?** À terme, nous souhaitons que ces travaux mènent à une implémentation sur de vrais robots et ne restent pas sur de la simulation. Une des questions qui se pose alors est de savoir à quel point l'IDSM est un modèle adapté pour cela. De nombreuses caractéristiques de l'IDSM nous intéressent et sont détaillées dans la section 3.5. Cependant, l'objectif initial de l'IDSM est l'encodage de comportements répétitifs assimilables à des habitudes, ce qui peut sembler à l'opposé d'un modèle cherchant à adapter les comportements initiaux (assimilables aux réflexes) d'un agent artificiel aux changements de son environnement pour augmenter sa capacité d'adaptation. Nous discutons ici de modifications futures qui pourraient être intéressantes de tester pour conserver l'IDSM ou *a minima* le côté carte sensorimotrice basée sur des nœuds.

Si un modèle d'apprentissage non-associatif est utilisé avec la création de nœuds pour mémoriser la modification du comportement provoquée par l'apprentissage, il faut aussi penser dans quelle mesure on souhaite rendre possible une future modification ou oubli d'un tel comportement acquis. En effet, l'IDSM original permet une sorte d'apprentissage en un coup ("*one-shot learning*"), où dès la première passe dans une zone de l'espace sensorimoteur peu dense en nœuds, l'agent laisse une trace de nœuds l'entraînant à toujours suivre cette même trajectoire sensorimotrice lors de futurs passages (sous réserve que les conditions sensorielles restent similaires à celles du premier passage. **En l'état**, les mises à jours des poids de l'IDSM original semblent avoir un impact de faible envergure sur la dynamique de l'espace sensorimoteur par rapport à l'ajout de nouveaux nœuds, impact qui nous semble insuffisant pour une modification significative du comportement acquis ou pour son oubli. Une piste est la modification des valeurs brutes utilisées dans l'équation 3.5 de l'IDSM original. Actuellement, la baisse de poids est de 1 par unité de temps, son renforcement est de 10 par unité de temps selon la distance à l'agent. Une seconde piste serait d'augmenter le paramètre  $k_w$ . Une troisième piste, qui peut être complémentaire, serait de rajouter un mécanisme de suppression (oubli) des nœuds pour limiter l'effet de "premier arrivé seul servi", par exemple s'il passe en deçà d'un certain poids et/ou si sa composante vitesse  $N_v$  n'est plus d'actualité avec la trajectoire de l'agent lorsqu'il repasse à proximité. De manière plus progressive, la composante  $N_v$  pourrait être modifiée par rapport à la trajectoire actuelle de l'agent qui repasserait à proximité, modification proportionnelle au poids  $N_w$  du nœud.

Augmenter la valeur du paramètre  $k_t$  pourrait aussi rendre le modèle plus plastique dans le temps et permettrait plus facilement de mettre à jour l'information en un point de l'espace sensorimoteur déjà visité. En effet,  $k_t$  sert de seuil de saturation de la densité  $\phi_N$  pour la création de nœuds. L'augmenter permettrait alors d'avoir plus facilement plusieurs nœuds en un même point

---

de l'espace sensorimoteur, avec des  $N_v$  potentiellement différents. Le cas actuel où un nœud "monopolise" un point de l'espace sensorimoteur avec son  $N_v$  pourrait avoir comme équivalence le cas extrême où ce point de l'espace serait saturé par plusieurs nœuds contenant la même information  $N_v$  (une sorte d'apprentissage par cœur). On peut imaginer aussi qu'avec un  $k_t$  important il y aurait des zones que l'on pourrait considérer "stables" avec des nœuds contenant des  $N_v$  identiques ou similaires (après passage de l'agent dans des conditions sensorimotrices similaires, dans le cas de passages répétés proches dans le temps), et des zones "instables" avec des nœuds contenant des  $N_v$  plus divergents (après un passage de l'agent dans des conditions différentes, passages peut-être plus espacés dans le temps). Comment de telles zones s'articuleraient avec nos modèles et les apprentissages non-associatifs n'est pas clair, mais peut-être pourraient-elles être une piste pour favoriser l'exploitation (à travers les zones stables) ou l'exploration (à travers les zones instables)? Dans quelle mesure un mécanisme permettant de renforcer de telles zones en fonction du niveau énergétique d'un robot (par exemple exploration quand énergie haute, exploitation sinon) empiéterait sur l'autonomie que nous voudrions qu'il acquiert?

Enfin, pour terminer cette réflexion sur l'IDSM, il n'est pas clair pour le moment de savoir quels comportements modifiés par les apprentissages non-associatifs auraient intérêt à être intégrés dans l'espace sensorimoteur à travers de nouveaux nœuds et lesquels devraient être oubliés plus rapidement et facilement avec les mécanismes de récupération, sans passer par des nœuds. Pourtant, il est important de réfléchir à la manière d'évaluer cette dualité rigidité contre plasticité, sur quels critères, cela en fonction des objectifs donnés à l'agent ou au robot. En particulier, les pistes suggérées ici pourraient, à l'inverse de l'IDSM original, rendre le modèle trop plastique à la moindre modification de l'environnement, empêchant alors l'agent d'avoir un comportement régulier.

**Réflexion sur l'autonomie de l'agent** Cette dernière réflexion est aussi l'une qui a servi de source à ces travaux de thèse. De quoi a besoin un agent artificiel pour être autonome? Dans quelle mesure ce que le développeur lui donne initialement contient déjà trop d'informations pour qu'il reste adaptatif, lui fournissant la réponse directement avant même le problème? Et d'ailleurs, avant tout, sur quoi évaluer l'autonomie d'un agent artificiel? Cette thèse ne répond pas à ces questions, qui restent bien entendu ouvertes.

Pour illustrer ces questions, prenons l'habituation. En le dotant d'un mécanisme d'habituation, nous disons explicitement à notre agent qu'un stimulus répété doit être ignoré. Par la suite, ne l'évaluer que dans une situation où il fait face à une répétition d'un stimulus ne donne pas d'informations sur une éventuelle autonomie.

En soi, nous pensons que les apprentissages non-associatifs ne sont pas suffisants pour qu'un agent artificiel soit autonome selon les tâches qu'il a à effectuer et selon les changements possibles de son environnement, mais qu'ils sont nécessaires pour un tel objectif, qu'ils sont l'une des briques de base avec lesquelles l'agent devrait être construit. Selon ce que l'on considère être de l'autonomie, doter l'agent de mécanismes d'apprentissages non-associatifs peut aussi être lui donner plus d'autonomie car cela permet de fonder des changements de comportements sur des mécanismes généraux plutôt que sur des instructions pré-câblées plus spécialisées.





# Bibliographie

- ALEX, J Yu et Catharine H RANKIN (2017). « Nonassociative Learning in Invertebrates ». In : *The Oxford Handbook of Invertebrate Neurobiology*. Oxford University Press.
- AVARGUÈS-WEBER, Aurore, Adrian G DYER et Martin GIURFA (2011). « Conceptualization of above and below relationships by an insect ». In : *Proceedings of the Royal Society B : Biological Sciences* 278.1707, p. 898-905.
- BARRANTES, Israel, Jeremy LEIPZIG et Wolfgang MARWAN (2012). « A next-generation sequencing approach to study the transcriptomic changes during the differentiation of *Physarum* at the single-cell level ». In : *Gene regulation and systems biology* 6, GRSB-S10224.
- BARRON, Andrew B et al. (2015). « Embracing multiple definitions of learning ». In : *Trends in neurosciences* 38.7, p. 405-407.
- BENHAMOU, Simon et Pierre BOVET (1989). « How animals use their environment : a new look at kinesis ». In : *Animal Behaviour* 38.3, p. 375-383.
- BI, Song et al. (2017). « On a neural network model based on non-associative learning mechanism and its application ». In : *2017 6th Data Driven Control and Learning Systems (DDCLS)*. IEEE, p. 679-684.
- BOISSEAU, Romain P, David VOGEL et Audrey DUSSUTOUR (2016). « Habituation in non-neural organisms : evidence from slime moulds ». In : *Proceedings of the Royal Society B : Biological Sciences* 283.1829, p. 20160446.
- BREAZEAL, Cynthia et Brian SCASSELLATI (1999). « A context-dependent attention system for a social robot ». In : *rn* 255.3.
- BURKHARDT, Richard W (1999). « Ethology, natural history, the life sciences, and the problem of place ». In : *Journal of the History of Biology* 32.3, p. 489-508.
- BYRNE, John H et Robert D HAWKINS (2015). « Nonassociative learning in invertebrates ». In : *Cold Spring Harbor perspectives in biology* 7.5, a021675.
- CAREW, Thomas J, Edgar T WALTERS et Eric R KANDEL (1981). « Classical conditioning in a simple withdrawal reflex in *Aplysia californica* ». In : *Journal of Neuroscience* 1.12, p. 1426-1437.

- DAMPER, Robert I, Richard LB FRENCH et Tom W SCUTT (2000). « ARBIB : An autonomous robot based on inspirations from biology ». In : *Robotics and Autonomous systems* 31.4, p. 247-274.
- DE HOUWER, Jan, Dermot BARNES-HOLMES et Agnes MOORS (2013). « What is learning? On the nature and merits of a functional definition of learning ». In : *Psychonomic Bulletin & Review* 20.4, p. 631-642.
- DOMJAN, Michael P (2014). *The principles of learning and behavior*. Cengage Learning.
- DUSSUTOUR, Audrey (2021). « Learning in single cell organisms ». In : *Biochemical and Biophysical Research Communications* 564, p. 92-102.
- EGBERT, Matthew (2018). « Investigations of an adaptive and autonomous sensorimotor individual ». In : *ALIFE 2018 : The 2018 Conference on Artificial Life*. MIT Press, p. 343-350.
- EGBERT, Matthew D et Xabier E BARANDIARAN (2014). « Modeling habits as self-sustaining patterns of sensorimotor behavior ». In : *Frontiers in human neuroscience* 8, p. 590.
- EISENSTEIN, Edward M et al. (2012). « Some new speculative ideas about the “behavioral homeostasis theory” as to how the simple learned behaviors of habituation and sensitization improve organism survival throughout phylogeny ». In : *Communicative & Integrative Biology* 5.3, p. 233-239.
- EISENSTEIN, EM, D EISENSTEIN et James C SMITH (2001). « The evolutionary significance of habituation and sensitization across phylogeny : A behavioral homeostasis model ». In : *Integrative Physiological & Behavioral Science* 36.4, p. 251-265.
- GILLARD, Tristan, Jérémy FIX et Alain DUTECH (2022). « [Re] Modeling habits as self-sustained patterns of sensorimotor behavior ». In : *The ReScience journal*.
- GRAY, Peter O (2011). *Psychology*. 6th. Worth Publishers.
- GROSS, Hans J et al. (2009). « Number-based visual generalisation in the honeybee ». In : *PloS one* 4.1, e4263.
- GROVES, Philip M, Diana LEE et Richard F THOMPSON (1969). « Effects of stimulus frequency and intensity on habituation and sensitization in acute spinal cat ». In : *Physiology & Behavior* 4.3, p. 383-388.
- GROVES, Philip M et Richard F THOMPSON (1970). « Habituation : a dual-process theory. » In : *Psychological review* 77.5, p. 419.
- HALL, Geoffrey et Gabriel RODRÍGUEZ (2020). « When the stimulus is predicted and what the stimulus predicts : Alternative accounts of habituation. » In : *Journal of Experimental Psychology : Animal Learning and Cognition* 46.3, p. 327.
- HARA, Sayuri et Jiro OHTA (1981). « Studies on cyst formation in *Physarum polycephalum myxamoebae* ». In : *The botanical magazine= Shokubutsu-gaku-zasshi* 94.2, p. 83-89.

- HAWKINS, Robert D et Eric R KANDEL (1984). « Is there a cell-biological alphabet for simple forms of learning? » In : *Psychological review* 91.3, p. 375.
- HAWKINS, Robert D. (1989). « A Biologically Based Computational Model for Several Simple Forms of Learning ». In : *Computational Models of Learning in Simple Neural Systems*. Sous la dir. de Robert D. HAWKINS et Gordon H. BOWER. T. 23. Psychology of Learning and Motivation. Academic Press, p. 65-108.
- HONG, Qinghui et al. (2020). « Memristive circuit implementation of biological nonassociative learning mechanism and its applications ». In : *IEEE Transactions on Biomedical Circuits and Systems* 14.5, p. 1036-1050.
- LACHMAN, Sheldon J (1997). « Learning is a process : Toward an improved definition of learning ». In : *The Journal of psychology* 131.5, p. 477-480.
- MARSLAND, S (1999). « A Model of Habituation Applied to Mobile Robots ». In : *Proc. TIMR, Towards Intelligent Mobile robots, Bristol, 1999*.
- MARSLAND, Stephen (2009). « Using habituation in machine learning ». In : *Neurobiology of learning and memory* 92.2, p. 260-266.
- MARSLAND, Stephen, Ulrich NEHMZOW, Jonathan SHAPIRO et al. (2000). « Detecting novel features of an environment using habituation ». In : *Proc. Simulation of Adaptive Behavior*. Citeseer.
- MARSLAND, Stephen, Jonathan SHAPIRO et Ulrich NEHMZOW (2002). « A self-organising network that grows when required ». In : *Neural networks* 15.8-9, p. 1041-1058.
- MCDIARMID, Troy A, Alex J YU et Catharine H RANKIN (2019). « Habituation is more than learning to ignore : multiple mechanisms serve to facilitate shifts in behavioral strategy ». In : *BioEssays* 41.9, p. 1900077.
- MCSWEENEY, Frances K et Eric S MURPHY (2009). « Sensitization and habituation regulate reinforcer effectiveness ». In : *Neurobiology of learning and memory* 92.2, p. 189-198.
- MÉDIONI, Jean et Marie-Cécile ROBERT (1969). « L'apprentissage chez les invertébrés ». In : *L'année psychologique* 69.1, p. 161-208.
- MOORE, Bruce R (2004). « The evolution of learning ». In : *Biological Reviews* 79.2, p. 301-335.
- MORENO, Carlos B et Jairo MUÑOZ-DELGADO (2007). « An account on the history of ethology ». In : *Suma Psicológica* 14.2, p. 213-224.
- NOVIANTO, Rony, Benjamin JOHNSTON et Mary-Anne WILLIAMS (2013). « Habituation and sensitisation learning in ASMO cognitive architecture ». In : *International Conference on Social Robotics*. Springer, p. 249-259.
- PERRY, Clint J, Andrew B BARRON et Ken CHENG (2013). « Invertebrate learning and cognition : relating phenomena to neural substrate ». In : *Wiley Interdisciplinary Reviews : Cognitive Science* 4.5, p. 561-582.

- PYKE, Graham H (2015). « Understanding movements of organisms : it's time to abandon the Lévy foraging hypothesis ». In : *Methods in Ecology and Evolution* 6.1, p. 1-16.
- RANDLETT, Owen et al. (2019). « Distributed plasticity drives visual habituation learning in larval zebrafish ». In : *Current Biology* 29.8, p. 1337-1345.
- RANKIN, Catharine H et al. (2009). « Habituation revisited : an updated and revised description of the behavioral characteristics of habituation ». In : *Neurobiology of learning and memory* 92.2, p. 135-138.
- RANKIN, Catherine H, Christine DO BECK et Catherine M CHIBA (1990). « *Caenorhabditis elegans* : a new model system for the study of learning and memory ». In : *Behavioural brain research* 37.1, p. 89-92.
- RESCORLA, Robert A (1972). « A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and nonreinforcement ». In : *Current research and theory*, p. 64-99.
- RICHARDS, Robert J (1987). *Darwin and the emergence of evolutionary theories of mind and behavior*. University of Chicago Press.
- RUESCH, Jonas et al. (2008). « Multimodal saliency-based bottom-up attention a framework for the humanoid robot icub ». In : *2008 IEEE International Conference on Robotics and Automation*. IEEE, p. 962-967.
- SHETTLEWORTH, Sara J (2001). « Animal cognition and animal behaviour ». In : *Animal behaviour* 61.2, p. 277-286.
- SKINNER, BF (1938). « The behavior of animals : An experimental analysis ». In : *New York : Appleton-Century Crofts*.
- SKINNER, Burrhus Frederic (1965). *Science and human behavior*. 92904. Simon et Schuster.
- (1984). « The evolution of behavior. » In : *Journal of the experimental analysis of behavior* 41.2, p. 217.
- SMITH-FERGUSON, Jules et Madeleine BEEKMAN (2019). « Who needs a brain? Slime moulds, behavioural ecology and minimal cognition ». In : *Adaptive Behavior*, p. 1059712319826537.
- STANLEY, James C (1976). « Computer simulation of a model of habituation ». In : *Nature* 261.5556, p. 146-148.
- TAYLER, CK et GS SAAYMAN (1973). « Imitative behaviour by Indian Ocean bottlenose dolphins (*Tursiops aduncus*) in captivity ». In : *Behaviour* 44.3-4, p. 286-298.
- THOMPSON, Richard F (2009). « Habituation : a history ». In : *Neurobiology of learning and memory* 92.2, p. 127.
- THOMPSON, Richard F et William A SPENCER (1966). « Habituation : a model phenomenon for the study of neuronal substrates of behavior. » In : *Psychological review* 73.1, p. 16.

- 
- THORNDIKE, Edward L (1898). « Animal intelligence : An experimental study of the associative processes in animals. » In : *The Psychological Review : Monograph Supplements* 2.4, p. i.
- TOLMAN, Edward Chace et Charles H HONZIK (1930). « " Insight" in rats. » In : *University of California Publications in Psychology*.
- URIBE-BAHAMONDE, Yercio E et al. (2019). « A quantitative account of the behavioral characteristics of habituation : The sometimes opponent processes model of stimulus processing ». In : *Frontiers in Psychology* 10, p. 504.
- VOGEL, Edgar H, Fernando P PONCE et Allan R WAGNER (2019). « The development and present status of the SOP model of associative learning ». In : *Quarterly Journal of Experimental Psychology* 72.2, p. 346-374.
- WAGNER, A (1981). « SOP : A model of automatic memory processing in animal behavior ». In : *Information processing in animals : Memory mechanisms*.
- WANG, DeLiang (1993). « A neural model of synaptic plasticity underlying short-term and long-term habituation ». In : *Adaptive Behavior* 2.2, p. 111-129.
- WANG, DeLiang et Chochun HSU (août 1990). « SLONN : A Simulation Language for Modeling of Neural Networks ». In : *Simulation* 55.2, p. 69-83. ISSN : 0037-5497.
- WENG, Juyang (2007). « On developmental mental architectures ». In : *Neurocomputing* 70.13-15, p. 2303-2323.
- WOOLFORD, Felix MG et Matthew D EGBERT (2020). « Behavioural variety of a node-based sensorimotor-to-motor map ». In : *Adaptive Behavior* 28.6, p. 425-440.
- WOSNIACK, Marina E et al. (2017). « The evolutionary origins of Lévy walk foraging ». In : *PLoS computational biology* 13.10, e1005774.
- ZARCO, Mario et Matthew EGBERT (2019). « Different forms of random motor activity scaffold the formation of different habits in a simulated robot ». In : *ALIFE 2019 : The 2019 Conference on Artificial Life*. MIT Press, p. 582-589.



## Résumé

Les apprentissages non-associatifs sont largement observés à travers la phylogénie et apparaissent fondamentaux pour l'adaptation et, ainsi, la survie des organismes vivants. Ce travail de thèse explore des mécanismes d'adaptation inspirés de ces apprentissages non-associatifs. Nous proposons trois modèles computationnels de l'habituation, trois modèles de la sensibilisation spécifique au site et un modèle de pseudo-conditionnement. Nous développons ces modèles dans le cadre du milieu sensorimoteur déformable itératif ("Iterant Deformable Sensorimotor Medium", IDSM), un modèle abstrait récemment développé de la formation du comportement sensorimoteur. Les caractéristiques des modèles présentés sont étudiées et analysées à la lumière de notre objectif à long terme, qui est d'étudier de nouveaux mécanismes d'apprentissage non supervisés pour des agents artificiels autonomes.

## Abstract

Non-associative learning is widely observed throughout phylogeny and appears to be fundamental for the adaptation and, thus, the survival of living organisms. This thesis explores adaptation mechanisms inspired by these non-associative learnings. We propose three computational models of habituation, three models of site-specific sensitization and one model of pseudo-conditioning. We develop these models within the framework of the Iterant Deformable Sensorimotor Medium (IDSM), a recently developed abstract model of sensorimotor behavior formation. The characteristics of the presented models are studied and analyzed in light of our long-term goal of investigating new unsupervised learning mechanisms for autonomous artificial agents.

