*Sujet de Thèse, Informatique*

# Auto-Organisation Multi-Echelle pour l'Emergence de Comportement Sensorimoteurs coordonnés

## 1 Contact, Lieu

Cette thèse sera supervisées par :
— Alain Dutech (alain.dutech@inria.f), HdR en Informatique
— Jeremy Fix (jeremy.fix@centralesupelec.fr), Docteur en Informatique

Administrativement, cette thèse se déroulera au sein de l'équipe Biscuit (LORIA). Physiquement, la thèse aura lieu dans les locaux du LORIA (Vandœuvre-les-nancy) et au sein du campus de CenttraleSupélec de Metz.

## 2 Contexte

Une question récurrente en sciences cognitives est de savoir comment un agent peut développer de nouvelles compétences sensorimotrices à partir de son expérience et de ses interactions en apprenant au sein d'un environnement ouvert. Des auteurs ont suggéré qu'un agent intelligent est doté de primitives sensorimotrices innées (par exemple, des central pattern generators) (Mussa-Ivaldi and Solla, 2004; Ijspeert, 2008). Ces primitives, ou réflexes, permettent à l'agent d'initier des interactions basiques avec leur environnement. Une question difficile est de comprendre comment l'agent peut s'appuyer sur ces interactions, en utilisant leur régularités mais aussi leurs imprévus, pour développer de nouvelles compétences et de nouveaux comportements, dans le but d'améliorer sa "maître" de son environnent.

Alors que l'intelligence artificielle s'est initialement focalisée sur les capacités de raisonnement abstraits, ce défi est au cœur du domaine plus récent de la cognition incarnée (Brooks, 1991; Nöe, 2004; Warren, 2006) qui insiste sur l'importance du couplage entre le cerveau, le corps et l'environnement dans lequel évolue l'agent. Cet environnement offre à l'agent des opportunités d'action mais, pour profiter de ces opportunités, l'agent est très dépendant de ses propres capacités motrices. Il y a ainsi une interdépendance entre l'environnement et le corps/Cerveau de l'agent et il doit donc exister des mécanismes qui permettent de tirer partir de ce couplage pour se développer.

Il existe un formalisme mathématique, l'apprentissage par renforcement, qui répond à cette problématique de comment un agent peut apprendre à se comporter dans un environnement inconnu (sans modèle de cet environnement) (Sutton and Barto, 2016). Cette théorie formalise les notions de récompense, politique d'action, fonction de valeur et procure ainsi des algorithmes permettant à un agent d'apprendre des comportements optimaux du point de vue des récompenses reçues. Les modèles issus de la théorie de l'apprentissage par renforcement sont généralement formulés avec une notion de temps discrétisé avec, comme granularité, les différentes interactions. Certains travaux récents permettent une extension au temps continu (Doya, 2000), mais il reste des difficultés et des questions en ce qui concerne la plausibilité biologique de ces modèles. Bien que la plausibilité biologique ne sont pas un élément primordial de cette thèse, il n'en reste pas moins que le fait que ces modèles ne soient pas biologiquement plausibles indique qu'il doit exister des solutions alternatives pour expliquer l'apprentissage "naturel" de comportement sensorimoteurs.

La littérature en neurosciences est riche en témoignages indiquant que plusieurs structures profondes du cerveau des primates, les ganglions de la base, sont impliqués dans l'apprentissage sensorimoteur guidé. Plusieurs facteurs contribuent à cet apprentissage. Depuis le travail séminal de (Schultz et al., 1997), on sait que des évènements récompensés imprévus déclenchent des réponses au niveau des neurones dopaminergiques de l'aire tegmentale ventrale, une structure dont on sait

qu'elle est en interaction forte avec les ganglions de la base. Plusieurs travaux insistent sur le lien entre ces structures et la théorie de l'apprentissage par renforcement (Niv, 2009). D'autres neurotransmetteurs guident aussi l'apprentissage (Doya, 2002). Des auteurs suggèrent qu'il peut être intéressant de guider l'apprentissage en utilisant une forme de curiosité qui module la façon dont un agent se confronte à son environnement de manière à développer des capacités sensorimotrice qui ne sont pas encore maîtrisées mais qui semblent apprenables (Kalplan and Oudeyer, 2007). En plus d'être une source d'inspiration pour la théorie de l'apprentissage par renforcement, les travaux en neurosciences indiquent que nos comportements pourraient se découper en deux grandes familles : les habitudes et les comportements dirigés par un but (Graybiel, 2005; Yin and Knowlton, 2006). Les comportements dirigés par un but, qui sont sensibles à la valeur du but et aux contingences entre l'action et son résultat, semblent pouvoir devenir des habitudes qui, moins flexibles, sont plus rapides à mettre en œuvre. Ces deux types de comportements et leur substrats neuronaux sont une piste à creuser pour tenter de doter un agent de la capacité d'agrandir son vocabulaire sensorimoteur.

# 3 Sujet

Au cours de cette thèse, nous voudrions travailler avec des modèles d'apprentissage en temps continu et compatibles avec le cadre général de l'apprentissage par renforcement. Nous voudrions explorer la façon dont ces modèles pourraient permettre à un agent de se construire des comportements plus abstraits et plus complexes.

Ainsi, par exemple, plusieurs travaux ont montrés que des modèles définis au sein de la Théorie des Systèmes Dynamiques (DST) (Kelso, 1995) peuvent exhiber des comportements sensorimoteurs simples (Beer, 1995; Spencer et al., 2011). Malgré ces succès, de nombreuses questions restent ouvertes, et ce sont justement ce genre de questions que nous voudrions explorer au cours de cette thèse.

Dès lors, voici une liste non exhaustive de questions et de défis qui pourraient être étudiés :

— alors que nous savons que les *central pattern generator* sont de bons candidats pour mettre en œuvre des primitive sensorimotrices simples, on est encore loin de comprendre comment des modèles distribués pourraient s'appuyer sur ce type de primitives pour définir des comportements à plus long terme.
— comment ces comportements plus abstraits et plus complexes pourraient-ils être intégrés au sein du modèles et s'interfacer avec les comportements plus simples ?
— comment le système pourrait-il apprendre des comportements encore plus abstraits de manière a amorcer et développer ses capacités sensorimotrices ?
— peut-on rendre opérante les théories de (Warren, 2006; Keijzer, 2001) qui promeuve l'intérêt et le rôle des comportement anticipatifs ?
— est-il possible de s'inspirer des travaux sur l'apprentissage par renforcement hiérarchique avec options (Sutton et al., 1999; Dietterich, 2000) dans un cadre où le temps est considéré comme continu ?

# 4 Profil

Le candidat sera ouvert à l'utilisation de techniques et de schémas de calculs non-conventionnels et bio-inspirés, avec un background en apprentissage automatique. Comme ce sujet de thèse prend ses racines en sciences cognitives, nous attendons du candidat qu'il soit au moins curieux de domaines comme la philosophie, la psychologie et la biologie. Cette thèse n'est pas seulement théorique et le candidat sera amené à développer et expérimenter ses modèles sur des plateformes robotiques simulées ou réelles. Ainsi, des compétences en programmation (`C++`, `Python`) et une familiarité avec le système `Unix` sont indispensables. De plus, une expérience avec le Robot Operating System (ROS) et des robots simulés (par exemple V-REP) ou réel sera un plus apprécié.

# Références


Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial intelligence*, 72(1–2) :173–215.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, 47 :139–159.

Dietterich, T. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research (JAIR)*, 13 :227–303.

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

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4–6) :495–506.

Graybiel, A. (2005). The basal ganglia : learning new tricks and loving it. *Current Opinion in Neurobiology*, (15) :638—644.

Ijspeert, A. (2008). Central pattern generators for locomotion control in animals and robots : a review. *Neural Networks*, 21(4) :642–653.

Kalplan, F. and Oudeyer, P. (2007). In the search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, 1(225–236).

Keijzer, F. (2001). *Representation and behavior*. MIT Press.

Kelso, J. S. (1995). *Dynamic Patterns*. MIT Press.

Mussa-Ivaldi, F. and Solla, S. (2004). Neural primitives for motion control. *IEEE Journal of Oceanic Engineering*, 29(3) :640–650.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3) :139–154.

Nöe, A. (2004). *Action in Perception*. MIT Press.

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

Spencer, J. P., Perone, S., and Buss, A. T. (2011). Twenty years and going strong : A dynamic systems revolution in motor and cognitive development. *Child Dev Perspect*, 5(4) :260–266.

Sutton, R. and Barto, A. (2016). *Reinforcement learning : An Introduction*. 2nd edition, in progress.

Sutton, R., Precup, D., and Singh, S. (1999). Between MDPs and semi-MDPs : A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, (112) :118–211.

Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, 113(2) :358–389.

Yin, H. and Knowlton, B. (2006). The role of the basal ganglia in habit formation. *Nature Reviews, Neuroscience*, (7) :464—476.