Apprentissage par Renforcement : Au delà des Processus Décisionnels de Markov.

(Vers la cognition incarnée)

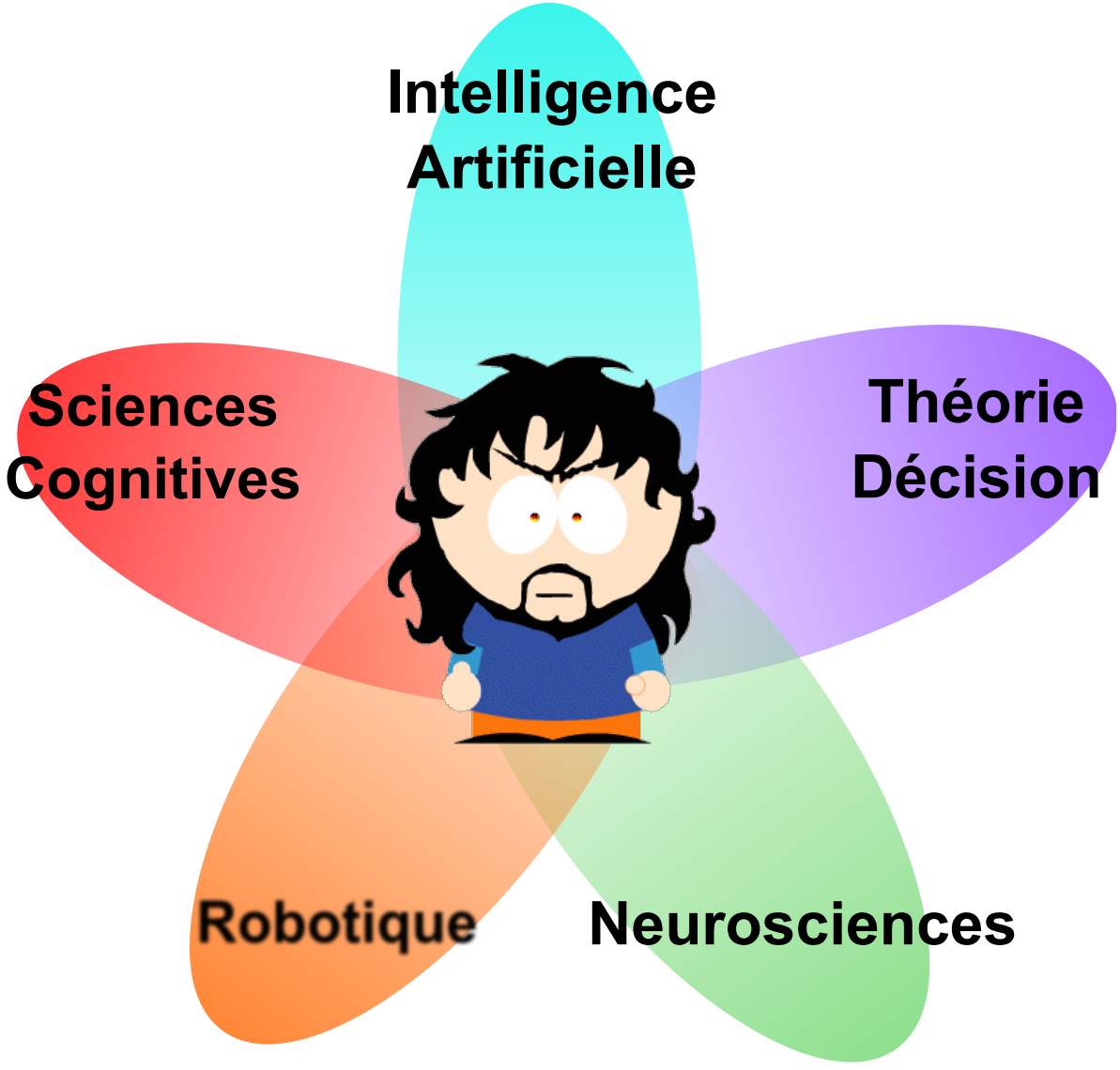
Alain Dutech

Equipe MAIA - LORIA - INRIA
Nancy, France
Web : http://maia.loria.fr
Mail : Alain.Dutech@loria.fr

2 déc. 2010

Positionnement

Quels sont les mécanismes de l'intelligence ?



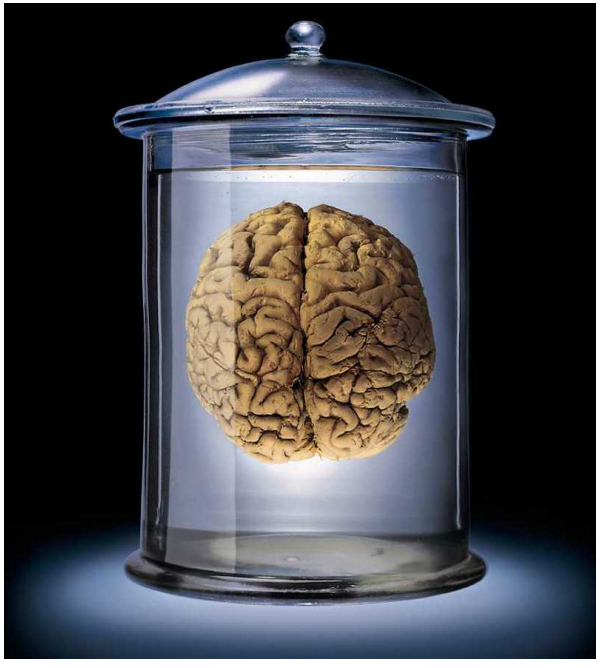
Plan de l'exposé

- Positionnement
- **Problématique**
- Retour vers le Passé
- Détails sur le Projet de Recherche
- Discutons

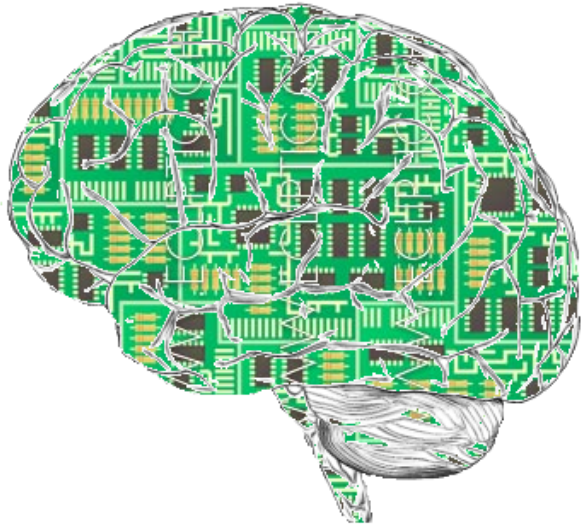
Intelligence Artificielle "**Classique**"

Cerveau = Machine à traiter des symboles

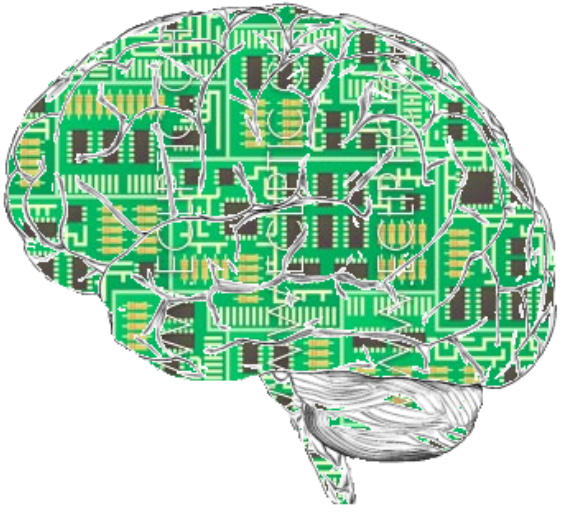
"Je ne saurais mieux me résumer qu'en disant qu'il existe désormais des machines capables de penser, d'apprendre et de créer."
[Newell et al., 1957]



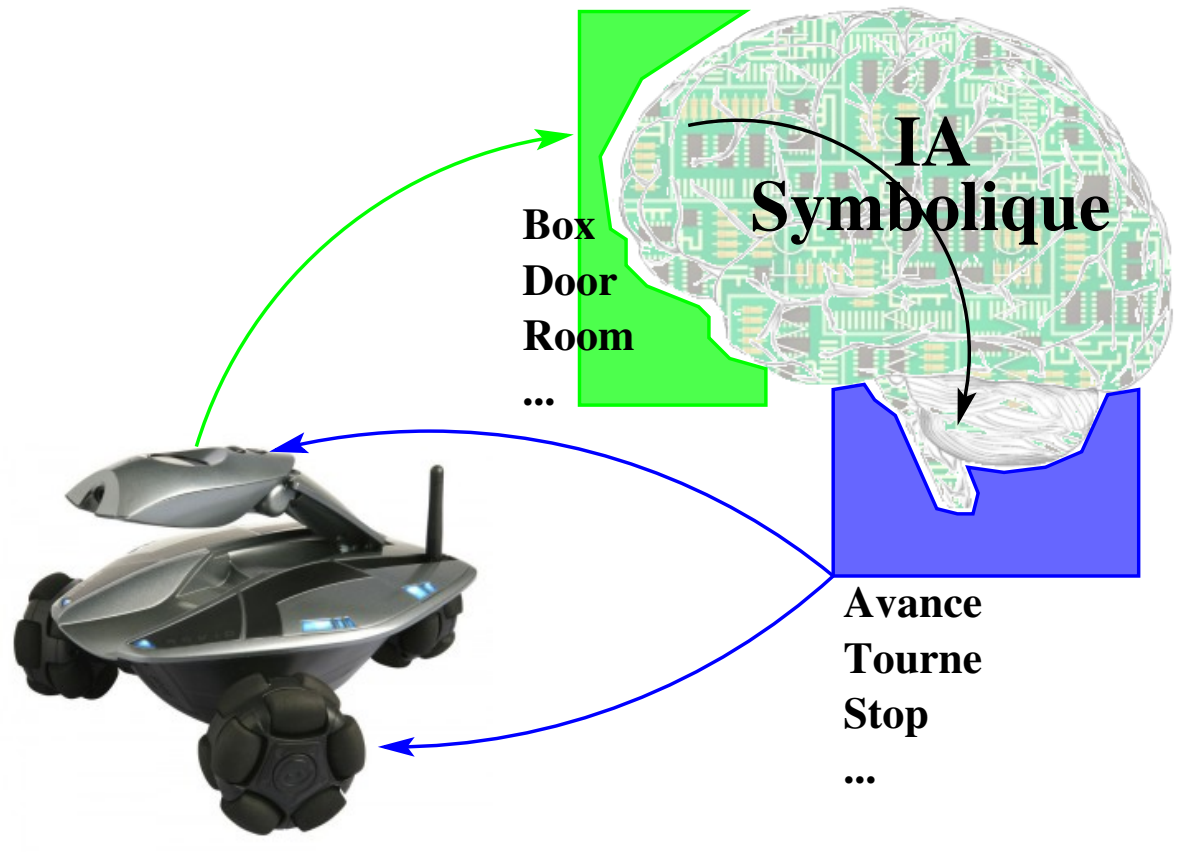
=



IA et “Ancrage des Symboles”

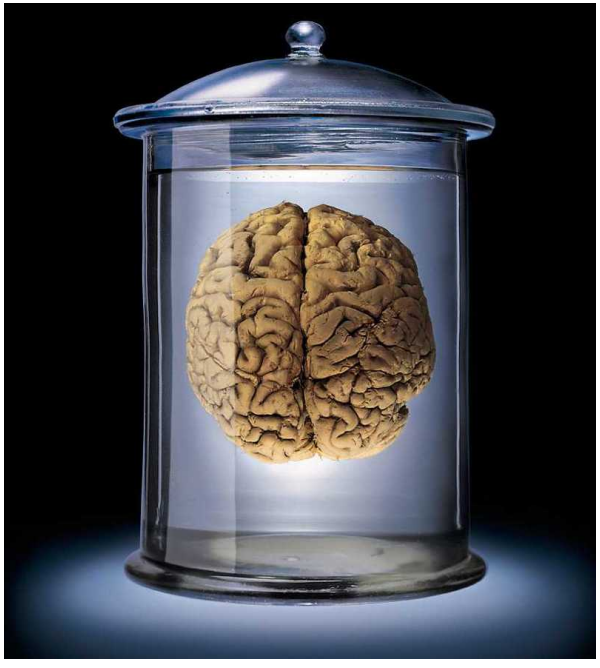


IA et "Ancrage des Symboles"

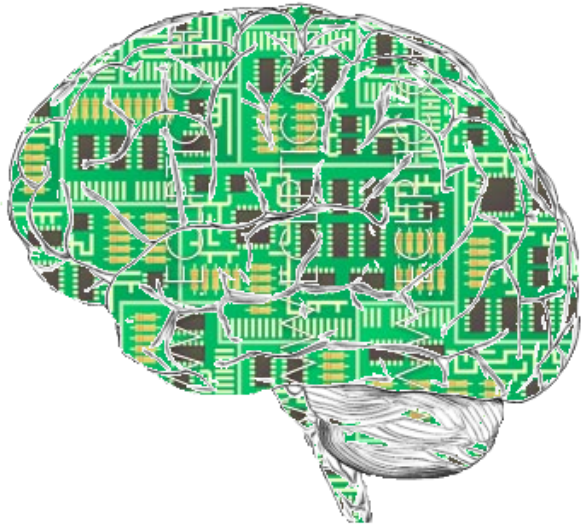


Intelligence Artificielle "**Classique**"

Cerveau = Machine à traiter des symboles



=



Problèmes Difficiles

- Ancrage des symboles [Harnad, 1990]
- Frame problem / Contexte pertinent [McCarthy and Hayes, 1969]
- Homonculus : intelligence dans l'intelligence [Gregory, 1987]

Cognition incarnée (“*Embodiment*”)

[Dreyfus, Brooks, Dennet, ...]

- Importance de corps+cerveau
- Rôle primordial des interactions (agent - environnement) motivées.
- Boucles sensori-motrices
- Pas de représentation abstraite *a priori*
- Mécanismes adaptation, apprentissage.

⇝ Emergence de “représentations” adaptées, situées, pertinentes

Grandes questions de la cognition incarnée

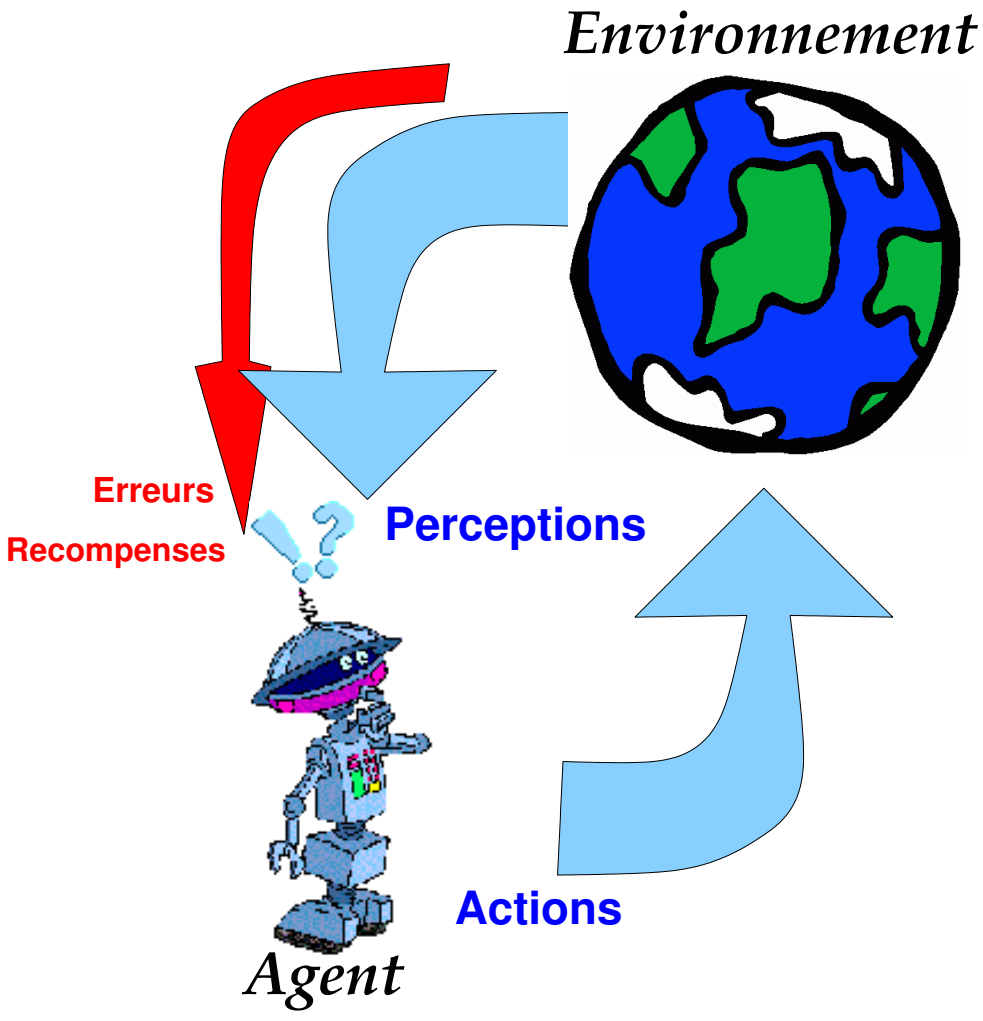
- Quelle place pour l'inné / l'acquis ?
- Quels mécanismes pour s'adapter, apprendre, décider ?
- Rôle et types des motivations ?
- Comment se créent les représentations ?

Une théorie essentiellement descriptive

Comment valider, prouver, étudier ?

$\rightsquigarrow$ Une démarche constructive. "Preuve" par l'exemple.

Une démarche constructive : Apprentissage par Renforcement



- Environnement incertain et inconnu
- Apprendre
- Agir optimalement en fonction du contexte
- Récompenses frustes

Propriétés de l'Apprentissage par Renforcement

[Puterman, 1994], [Sutton and Barto, 1998], [Groupe PDMIA, 2008], ...

Cadre formel : MDP

$\mathcal{S}$, ens. d'états ; $\mathcal{A}$, ens. d'actions ;
$p(s'|a,s)$, transitions ; $r(s,a)$, récompenses
$\rightsquigarrow$ Peut **apprendre** une politique optimale au sens de $\sum \gamma^t r_t$

Modélisation

- Large inspiration (neuro)-biologique
- Différents niveaux de modélisation

Assise théorique

- $\mathcal{S}$, $\mathcal{A}$ donnés *a priori*
- Récompenses *extérieure*
- **Propriété de Markov**. L'état doit contenir assez d'information pour prédire le futur

Non-Markovien ≈ élaboration représentations

[Aström, 1965], [Sondik, 1971], ...

Processus Partiellement Observable

$$\Pr(o_{t+1}|o_t, a_t) \neq \Pr(o_{t+1}|o_t, a_t, \cdots, a_1, o_0)$$

Comment se ramener à un processus **Markovien** ?
$\rightsquigarrow$ utiliser un espace d'états approprié (état d'information complet).

Agent doit se construire une représentation appropriée

Résumons-nous avant de plonger dans le passé

- Quels sont les mécanismes de la cognition ?

- Cognition incarnée : émergence par interactions motivées
- Validation par l'exemple

- Apprentissage par renforcement comme approche constructive

Problématique

Etudier l'élaboration de représentations pertinentes par le biais de l'apprentissage de comportements optimaux dans des Processus de Décision **Non-Markoviens**.

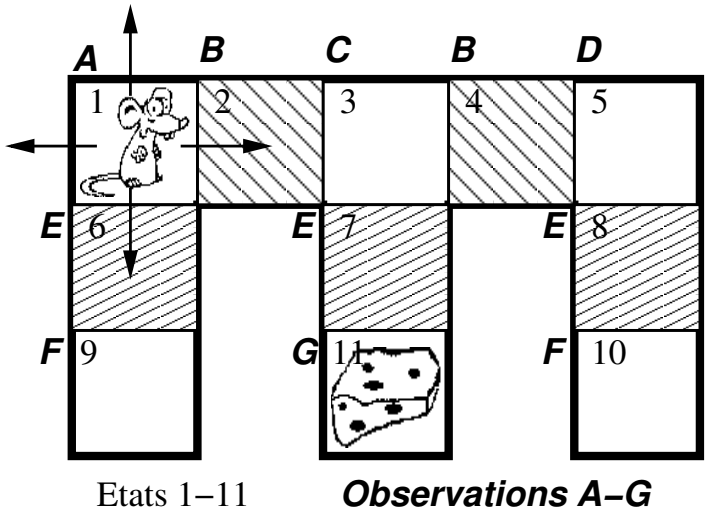
Plan de l'exposé

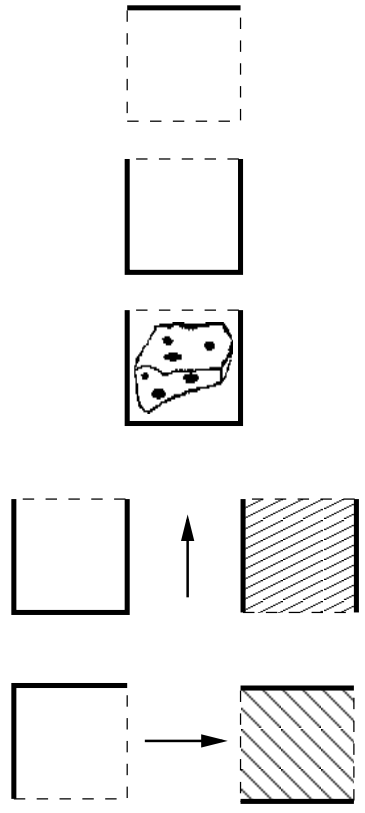
- Positionnement
- Problématique
- **Retour vers le passé**
- Détails sur le Projet de Recherche
- Discutons

Apprentissage et POMDP

Extension d'état "sélective" vs *belief-state*, *PSR*.

- extension sélective par variance de $Q(s, a)$ [Dutech, 2000]
- observation pertinente selon information mutuelle [Dutech and Scherrer, 2001]





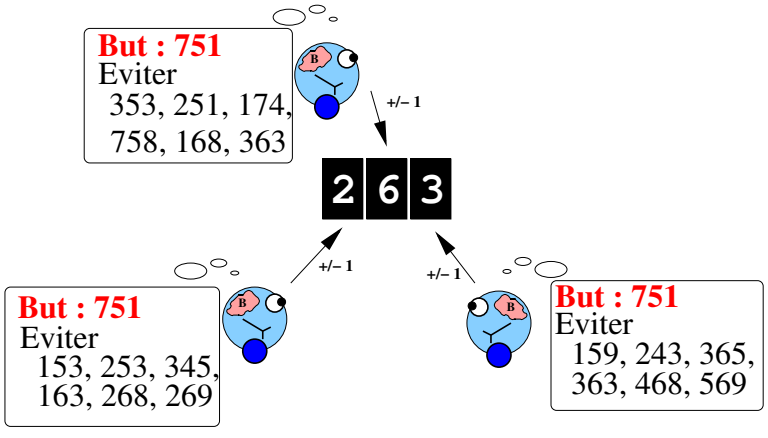
Confirmation de problèmes connus

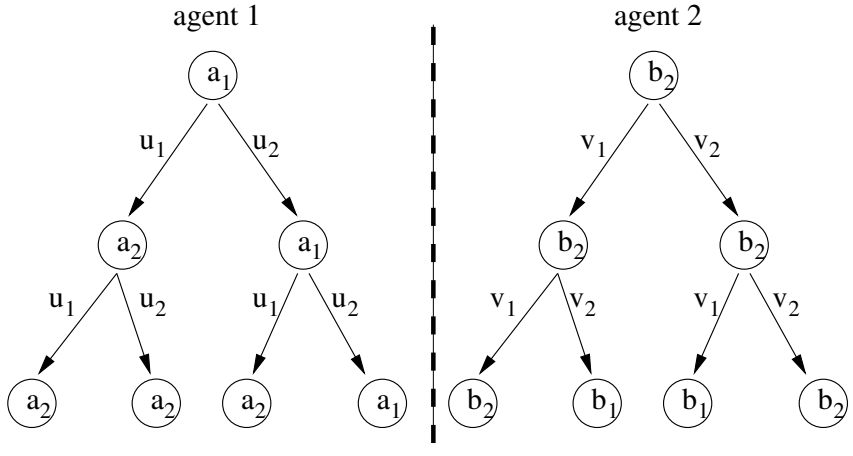
Limite des approches "comptables", "statistiques", "désincarnées" (explosion combinatoire)

Interactions inter-agents et Dec-POMDP

Motivé par guide, expert, imitation... **MAIS** nécessité d'un couplage [Hart and Mas-Colell, 2003]

- Quel type et quantité de couplage *a priori* ? [Aras et al., 2005]
- Complexité démesurée. Exploiter la structure des solutions [Aras and Dutech, 2010]





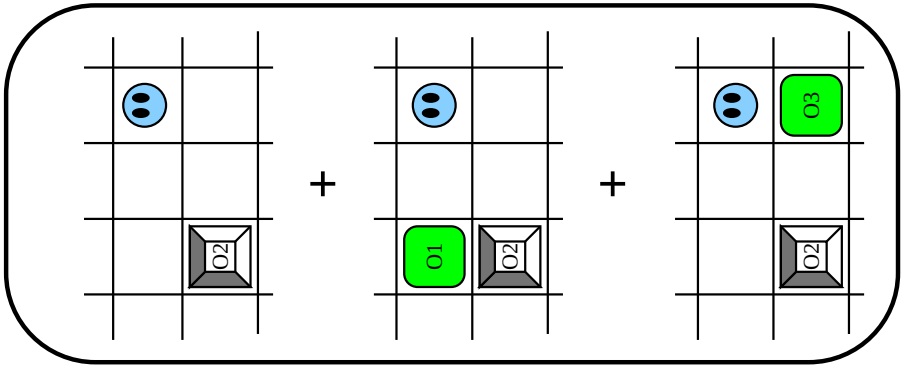
Difficulés supplémentaires non prioritaires

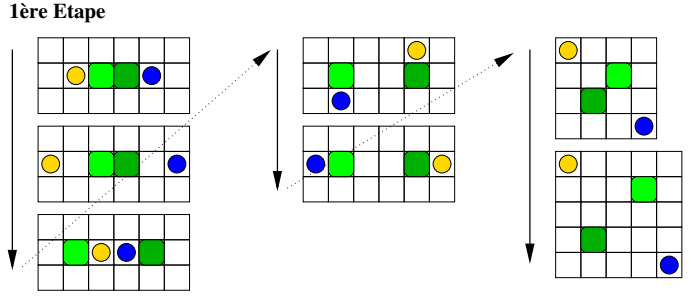
Capacités de couplage "innées" (modèle, motivation, communication) nécessaires, mais complexes et difficiles, pour améliorer l'apprentissage.

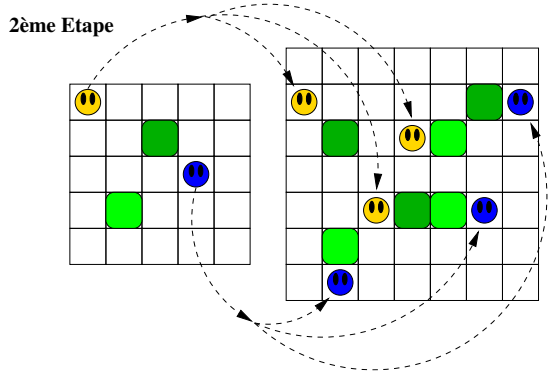
Piste 1 : approches incrémentales (non-optimales)

Guider/aider l'agent dans son apprentissage de tâches difficiles

- Construction incrémentale de comportements composés [Buffet et al., 2004]
- "*Shaping*" en tâche et en nombre [Buffet et al., 2007]





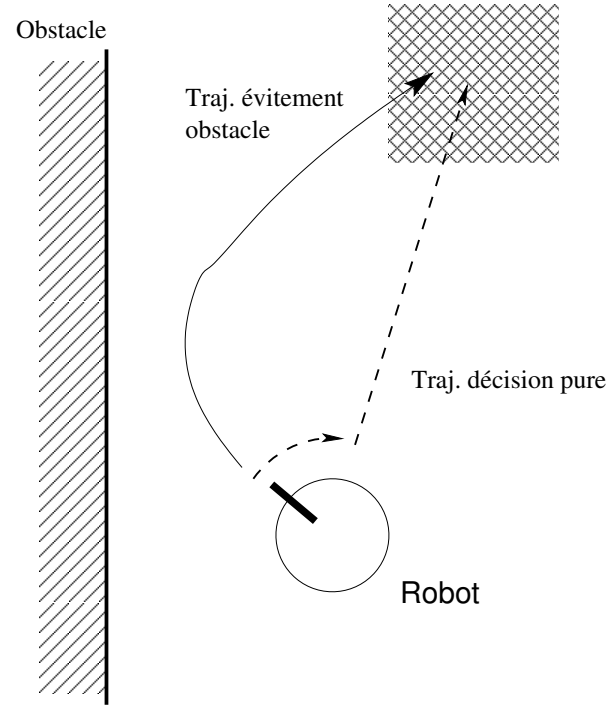


Premier pas d'une Approche Incrémentale

Faciliter l'apprentissage en adaptant/façonnant agent et environnement au fur et à mesure de l'amélioration des performances.

Piste 2 : Incarnation et robotique

- Richesse environnement non-contrôlé.
- Richesse des interactions possibles [Beaufort, 2009]



Apports de l'incarnation

Importance d'un robot pour interagir avec un environnement riche et non-contrôlé, dans des boucles sensori-motrices.

Plan de l'exposé

- Positionnement
- Problématique
- Retour vers le Passé
- **Détails sur le Projet de Recherche**
- Discutons

Projet de Recherche

Apprentissage par Renforcement incrémental, holistique et motivationnel

Problématique

Etude de l'élaboration de représentations pour des problématiques non-Markoviennes par le biais principal de l'Apprentissage par Renforcement

- Incrémental. Notamment dans des dimensions sensorielles et motrices [Lungarella et al., 2003]
- Holistique. Incarnation, apprentissage seul n'est pas suffisant. [Edelman/Sporns], [Gaussier]
- Motivationnel. Origines, rôles et effets du renforcement. [Oudeyer], [Redgrave et al., 2008]

Projet de Recherche

Apprentissage par Renforcement incrémental, holistique et motivationnel

Problématique

Etude de l'élaboration de représentations pour des problématiques non-Markoviennes par le biais principal de l'Apprentissage par Renforcement

- Incrémental. Notamment dans des dimensions sensorielles et motrices [Lungarella et al., 2003]
- Holistique. Incarnation, apprentissage seul n'est pas suffisant. [Edelman/Sporns], [Gaussier]
- Motivationnel. Origines, rôles et effets du renforcement. [Oudeyer], [Redgrave et al., 2008]

Projet décliné en 3 axes.

Axe 1 : Robotique cognitive développementale

Holistique, Incrémental

Apprendre par Renforcement avec un Robot
$\rightsquigarrow$ représentations, env. continu, efficacité.

- Quelles capacités innées pour le robot ?
- Sous quelle forme manipuler les données sensori-motrices ?
- Comment exploiter au mieux chaque interaction ?
- Comment "accompagner" l'apprentissage ?

Complémentaire

Motivation [Oudeyer], Act.continue [Sigaud],
Arch. Neuronale [edelman/Sporns, Gaussier], *etc.*

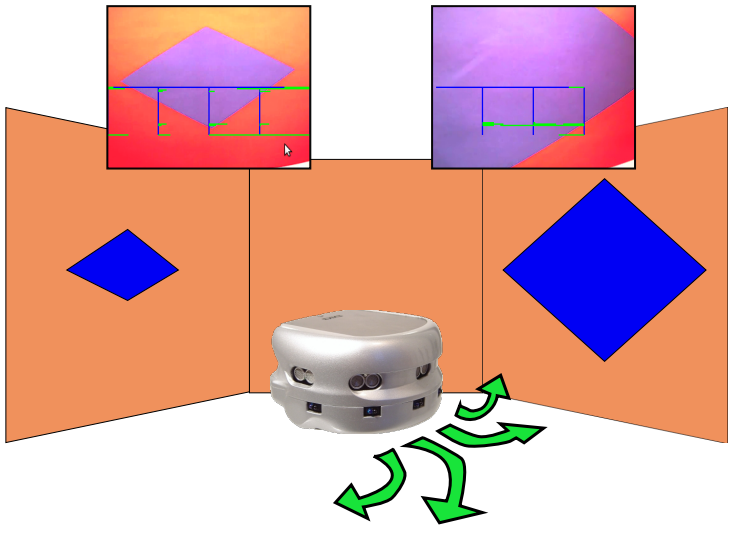
Axe 1 : Robotique cognitive développementale

Holistique, Incrémental

Apprendre par Renforcement avec un Robot

$\rightsquigarrow$ représentations, env. continu, efficacité.

- Quelles capacités innées pour le robot ?
- Sous quelle forme manipuler les données sensori-motrices ?
- Comment exploiter au mieux chaque interaction ?
- Comment "accompagner" l'apprentissage ?



[Sarzyniec, 2010]

Complémentaire

Motivation [Oudeyer], Act.continue [Sigaud], Arch. Neuronale [edelman/Sporns, Gaussier], *etc.*

Axe 2 : Neurosciences Computationnelles

Holistique, Motivationnel

Inspiration et meilleure compréhension du vivant

⇝ Collaborations, Bibliographie

⇝ Boucle Cortex-Ganglions de la Base-Thalamus

⇝ Neuro-modulation, Dopamine

⇝ [Sutton, Dayan, Schultz, Redgrave, Gurney, Girard, ...]

- Quelles représentations dans les boucles sensori-motrices ?
- Neuromodulateurs : origine ? effet ? rôle ?
- Existence et modèles d'une fonction d'apprentissage par renforcement ?
- Existence et modèles d'un Apprentissage synaptique neuromodulé ?

Axe 2 : Neurosciences Computationnelles

Holistique, Motivationnel

Inspiration et meilleure compréhension du vivant

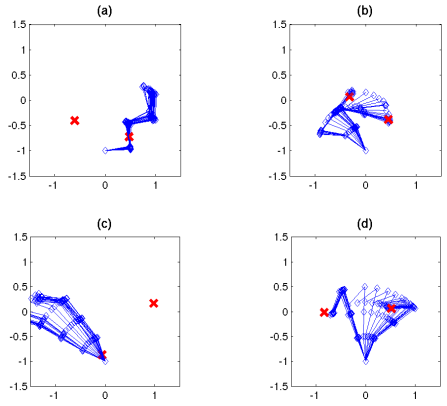
⇝ Collaborations, Bibliographie

⇝ Boucle Cortex-Ganglions de la Base-Thalamus

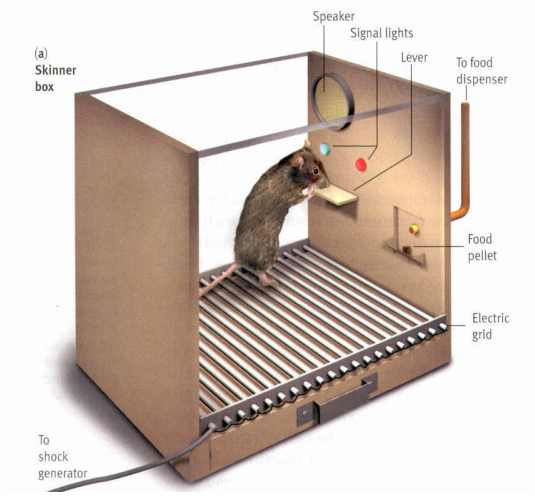
⇝ Neuro-modulation, Dopamine

⇝ [Sutton, Dayan, Schultz, Redgrave, Gurney, Girard, ...]

- Quelles représentations dans les boucles sensori-motrices ?
- Neuromodulateurs : origine ? effet ? rôle ?
- Existence et modèles d'une fonction d'apprentissage par renforcement ?
- Existence et modèles d'un Apprentissage synaptique neuromodulé ?



[Daucé and Dutech, 2010]



[Dutech et al., 2010]

Axe 3 : Apprentissage Auto-organisé dans SMA

Holistique, Motivationnel

Apprentissage + Emergence : au centre de l'élaboration de représentations

$\rightsquigarrow$ apprentissage par renforcement distribué $\rightsquigarrow$ auto-organisation

- Quel renforcement pour un Apprentissage distribué ?
- Effet de l'apprentissage sur l'Auto-organisation ?
- Améliorer la caractérisation de l'Auto-organisation ?
- Une ou des émergences ?

Complémentaire

Robustesse [Fates], App. Renf [Meuleau], Mesure [Shalizi], *etc.*

Mais pas tout seul !

Inspirations, tendances, complémentarités

- Rob. Développementale. Oudeyer, Edelman/Sporns, Sigaud, Sutton, Peters, Riedmiller, Gaussier, *etc.*
- NeuroSciences. Redgrave, Gurney, Dominey, Girard, Schultz, Hikosaka, Haber, Boraud, *etc.*
- Auto-organisation. Theraulaz, Dorigo, Deneubourg, Shalizi, *etc.*

Coopérations

$\rightsquigarrow$ EPI Cortex, ANR MAPS, NeuroInformatique C.N.R.S., A. Marchand, E. Daucé, à suivre...

Plan de l'exposé

- Positionnement
- Problématique
- Retour vers le Passé
- Détails sur le Projet de Recherche
- **Discutons**

Conclusion provisoire

Projet

Mécanismes d'Apprentissage par Renforcement Holistiques, Incrémentaux et Motivationnels pour des tâches non-Markoviennes en Robotique.

- Participer à la compréhension et création de mécanismes cognitifs.
- **Projet** vaste, ambitieux, à large spectre
- Beaucoup de questions, pistes incertaines

Conclusion provisoire

Projet

Mécanismes d'Apprentissage par Renforcement Holistiques, Incrémentaux et Motivationnels pour des tâches non-Markoviennes en Robotique.

- Participer à la compréhension et création de mécanismes cognitifs.
- **Projet** vaste, ambitieux, à large spectre
- Beaucoup de questions, pistes incertaines
- Dans un paysage scientifique défavorable...
  - Visibilité/Rayonnement/Concurrence vs Science
  - Projets, Innovation, ... , Productivité, Comptabilité
  - Formation des jeunes chercheurs

Références I

Aras, R. and Dutech, A. (2010).

An investigation into mathematical programming for finite horizon decentralized POMDPs.

Journal of Artificial Intelligence Research (JAIR), 37 :329–396.

Aras, R., Dutech, A., and Charpillet, F. (2005).

Cooperation in stochastic games through communication.

In *Proc. of the fourth Int. Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS'05)*, Utrech, Nederlands.

Aström, K. (1965).

Optimal control of Markov decision processes with incomplete state estimation.

Journal of Mathematical Analysis and Applications, 10 :174–205.

Références II

Beaufort, N. (2009).

Apprentissage optimiste et planification partielle pour un robot mobile.

Master's thesis, Université Henri Poincaré, Nancy I.

Buffet, O., Dutech, A., and Charpillet, F. (2004).

Self-growth of basic behaviors in an action selection based agent.

In *From Animals to Animats. Proc. of Int. Conf. on Simulation of Adaptive Behavior (SAB'04)*, Los Angeles, USA.

Buffet, O., Dutech, A., and Charpillet, F. (2007).

Shaping multi-agent systems with gradient reinforcement learning.

Autonomous Agent and Multi-Agent System Journal (AAMASJ), 15(2) :197–220.

Références III

Daucé, E. and Dutech, A. (2010).

Online Learning with Noise : A Kernel-Based Policy-Gradient Approach.

In *Conférence Française de Neurosciences Computationnelles - NeuroComp 2010*, Lyon France.

Dutech, A. (2000).

Solving POMDP using selected past-events.

In *Proceedings of the 14th European Conference on Artificial Intelligence, ECAI2000.*

Dutech, A., Coutureau, E., and Marchand, A. (2010).

Reinforcement Learning Approaches to Instrumental Contingency Degradation in Rats.

In *Conférence Française de Neurosciences Computationnelles - NeuroComp 2010*, Lyon France.

Références IV

Dutech, A. and Scherrer, B. (2001).

Learning to use contextual information for solving partially observable Markov decision problems.

In *Fifth European Workshop on Reinforcement Learning, EWRL-5*, Utrecht, Nederlands.

Gregory, R. (1987).

The Oxford companion to the mind.

Oxford University Press, Oxford, UK.

Groupe PDMIA (2008).

Processus Décisionnels de Markov en Intelligence Artificielle. (Edité par Olivier Buffet et Olivier Sigaud), volume 1 & 2.

Lavoisier - Hermes Science Publications.

Références V


Harnad, S. (1990).
The symbol grounding problem.
Physica D, 42 :335–346.


Hart, S. and Mas-Colell, A. (2003).
Uncoupled dynamics do not lead to Nash equilibrium.
American Economic Review, pages 1830–1836.


Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003).
Developmental robotics : a survey.
Connection Science, 15(4) :151–190.


McCarthy, J. and Hayes, P. (1969).
Some philosophical problems from the standpoint of artificial intelligence.
Machine Intelligence, 4 :463–502.

Références VI


Newell, A., Simon, H., and Shaw, J. (1957).
Preliminary report of general problem solving, (working paper 7).
Technical report, Carnegie Insitute of Technology.


Puterman, M. (1994).
Markov Decision Processes : discrete stochastic dynamic programming.
John Wiley & Sons, Inc. New York, NY.


Redgrave, P., Gurney, K., and Reynolds, J. (2008).
What is reinforced by phasic dopamine signals ?
Brain research reviews, 58(2) :322–339.


Sarzyniec, L. (2010).
Apprentissage par renforcement développemental pour la robotique autonome.
Master's thesis, Université Henri Poincaré, Nancy I.

Références VII

Sondik, E. (1971).

The optimal control of partially observable markov decision processes.

PhD thesis, Stanford University, California.

Sutton, R. and Barto, A. (1998).

Reinforcement Learning.

Bradford Book, MIT Press, Cambridge, MA.