

# Proposition d'un formalisme pour la construction automatique d'interactions dans les systèmes multi-agents réactifs

*Soutenance de thèse - Doctorat Informatique*

Vincent Thomas

Joël Quinqueton

*Rapporteur*

Philippe Mathieu

*Rapporteur*

Olivier Sigaud

*Examineur*

Didier Galmiche

*Examineur*

Christine Bourjot

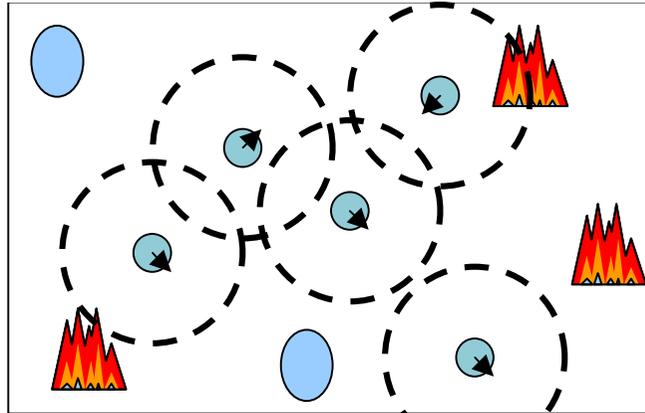
*Directeur*

Vincent Chevrier

*Directeur*

# Contexte scientifique

- Objectif à long terme
  - Concevoir des systèmes multi-agents réactifs pour la résolution collective de problèmes



- Sous-Problématique de l'intelligence artificielle:
  - IA incarnée, distribuée, rationnelle et réactive
- Problématique énoncée par Ferber [Ferber 97]:
  - Concevoir un système multi-agents consiste à répondre aux questions :
    - Quelle architecture donner aux agents ?
    - Quelles formes d'interaction introduire dans le système ?
    - Quels comportements donner aux agents et comment les adapter ?
    - Comment implanter ces systèmes ?

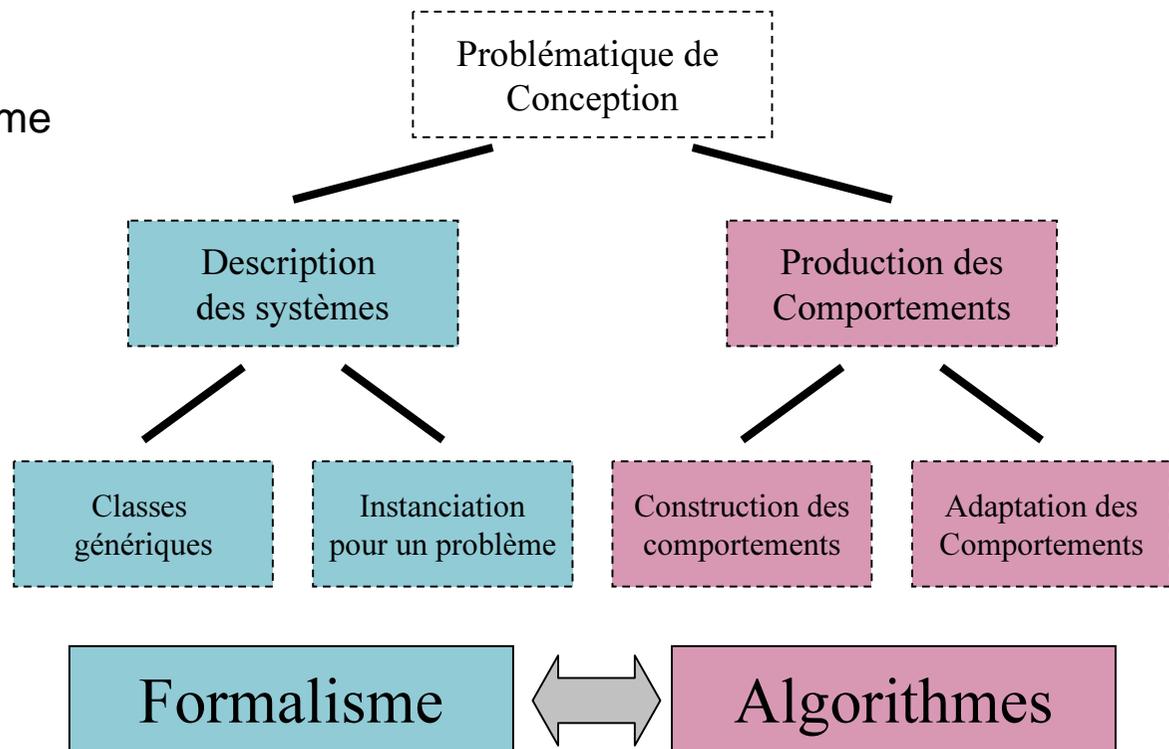
# Notre approche

- Contraintes de notre approche

- Perception partielle → Adaptative
- Sans connaissance a priori → Automatique
- Réutilisation → Générique

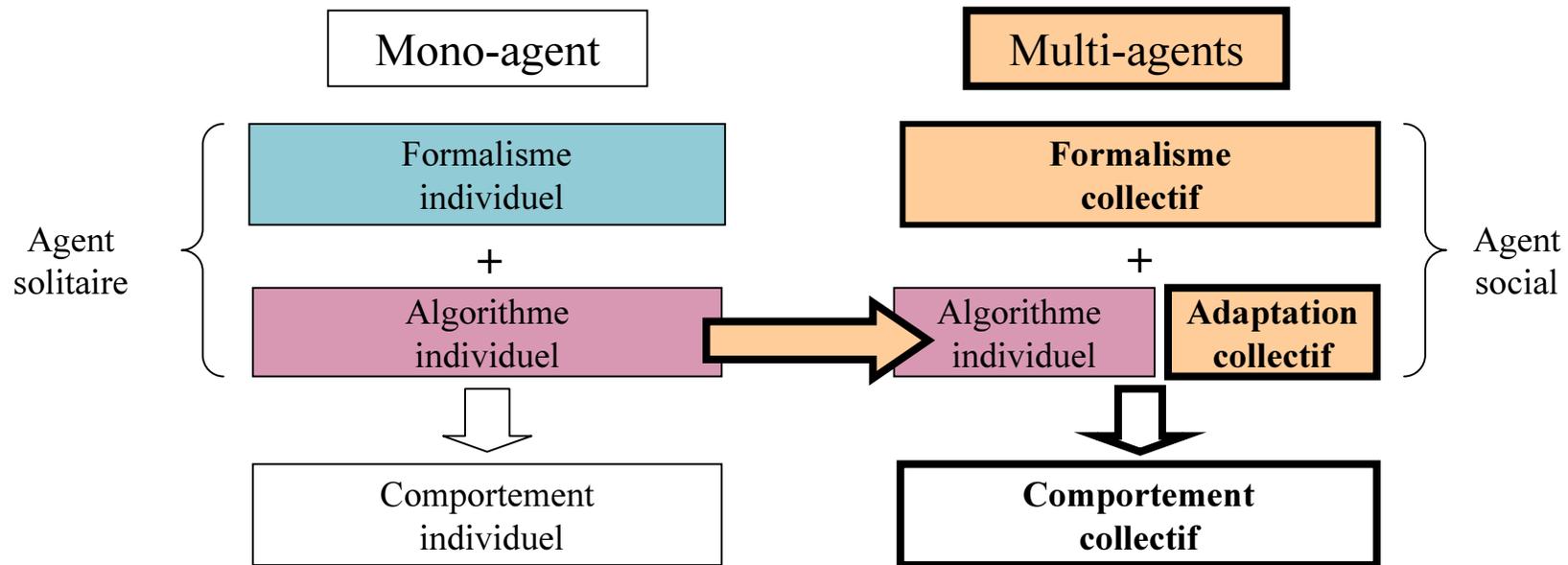
- Approches formelles

- Décomposition du problème



# Proposition

- Hypothèses de travail
  - Le collectif n'est pas forcément très éloigné de l'individuel
    - Global impossible
    - Individuel insuffisant
  - Réutiliser les techniques mono-agent existantes est pertinent
- Démarche : Construire agents sociaux
  - Qui considèrent la présence d'autres agents → Formalisme
  - Qui adaptent leurs comportements aux autres → Algorithme



# Plan

- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

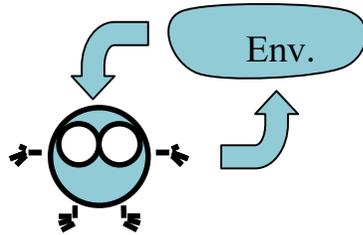
# Plan

- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

Quelle est la différence entre des systèmes mono-agent et des systèmes collectifs ?

# Mise en regard des concepts

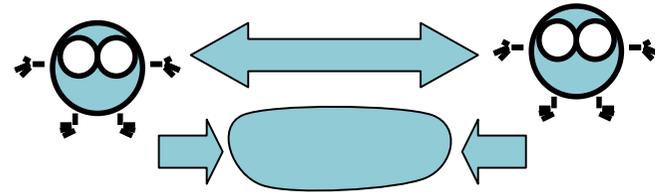
- Agent autonome réactif



- Entité [Russel Norvig95]
  - Percevant l'environnement
  - Agissant sur l'environnement
- Intelligence → Rationalité
- Conception
  - Description
    - Architecture
      - Effecteurs, Senseurs
      - Mémoire
  - Comportement
    - Fonction décision

Centré rationalité

- Système Multi-agents

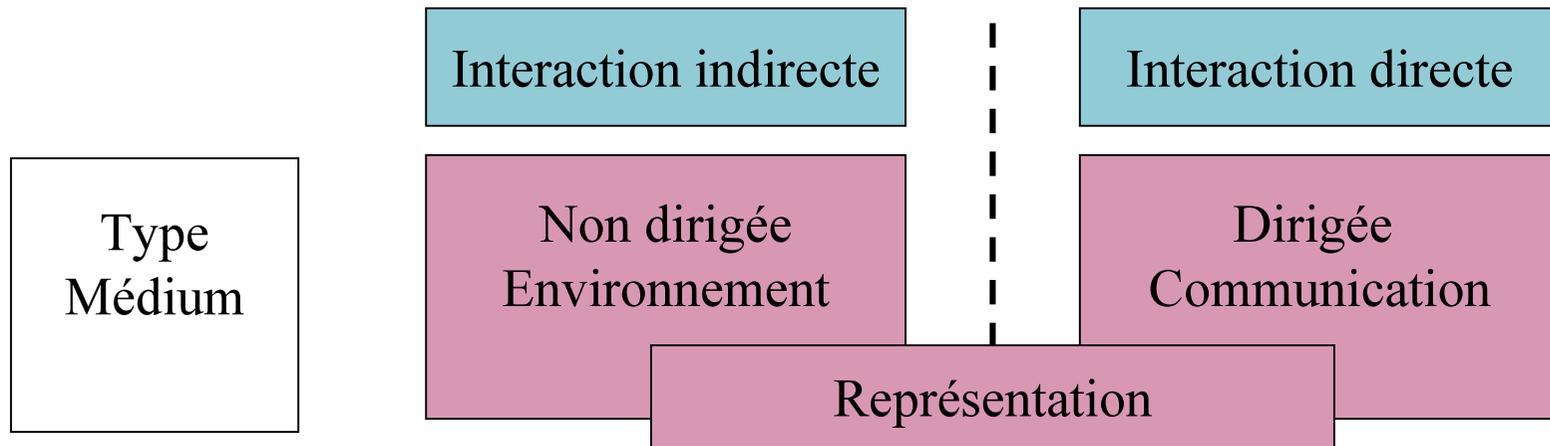


- Approche Voyelle [Demazeau95]  
**A** : Agents      **E**: Environnement,  
**I**: Interactions    **O**: Organisation
- Intelligence → Intelligence collective
- Conception
  - Description
    - Architecture des agents
    - Formes d'interaction
  - Comportement
    - Fonction décision
    - Adaptation aux autres

Centré interaction

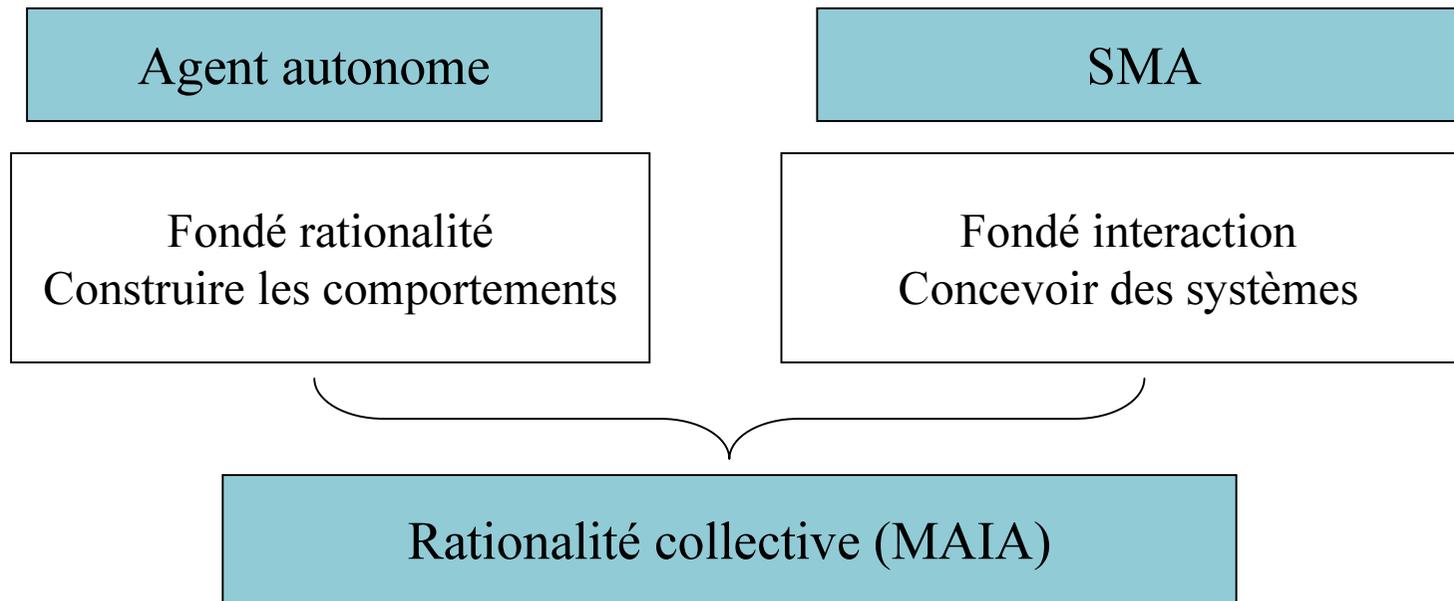
# Visions de l'interaction

- 3 visions de l'interaction dans la littérature
  - L'interaction comme situation
    - Attitude des agents les uns par rapport aux autres
    - → Systèmes coopératifs
  - L'interaction comme couplage
    - Influence mutuelle réciproque
    - → Essence des systèmes multi-agents
  - L'interaction comme moyen
    - Formes d'interaction proposées dans un système



# Bilan sur les concepts

- Deux communautés



## Thèse soutenue

La représentation explicite de l'interaction dans les formalismes permet aux agents de considérer les autres agents et d'apprendre localement à adapter leur comportement

# Plan

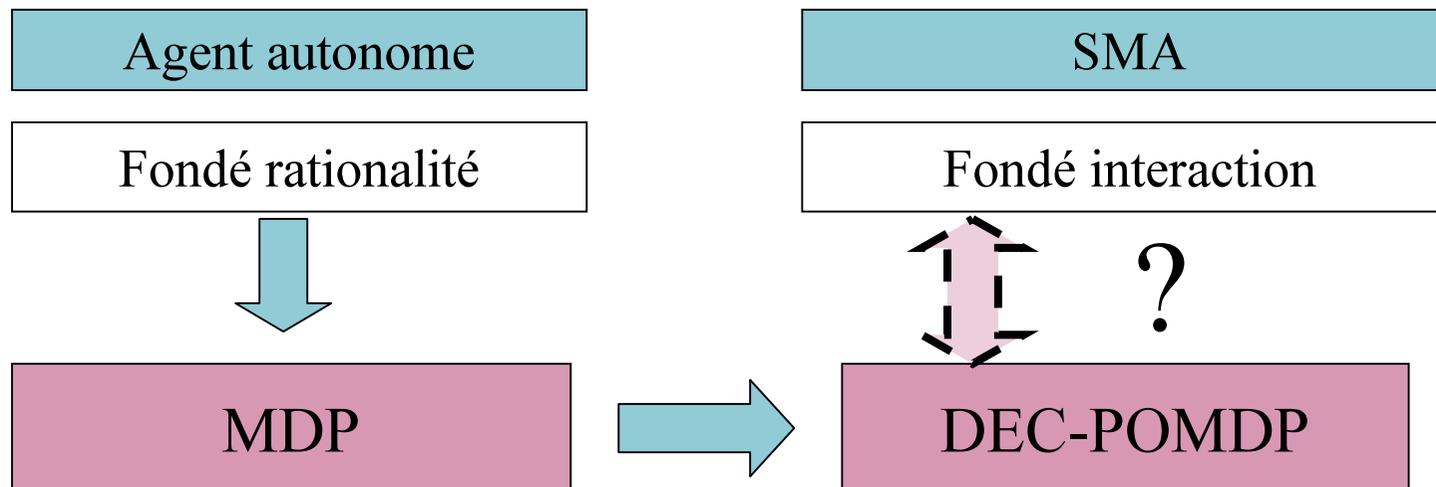
- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

Quelles formes d'interactions  
sont formalisées ?

# Modèles markoviens

- Intérêts des modèles markoviens
  - Problématique de construction de comportements
  - Réponses dans un cadre mono-agent
  - Existence d'extensions multi-agents

- Question:

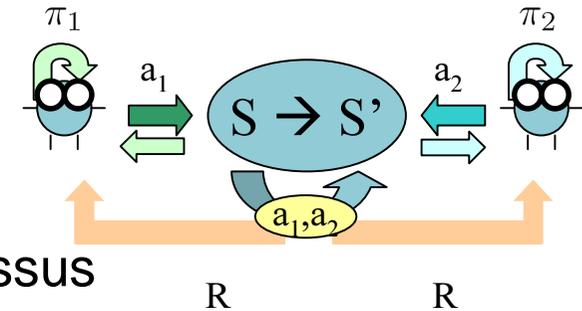


# Extension des formalismes MDP

- DEC-POMDP (Decentralized Partially Observable Markov Decision Process)

- Représenter agents réactifs

- Observabilité partielle
- Exécution décentralisée et simultanée



- Représenter problème sous forme d'un processus

- Matrice de transition  $T : S \times A_1 \times A_2 \times A_3 \times \dots \rightarrow S$
- Fonction de récompense  $R : S \times A_1 \times A_2 \times A_3 \times \dots \rightarrow P(\text{Re})$
- Actions des agents vues comme influences sur processus

- Problème de décision : Trouver les politiques locales des agents
  - pour maximiser la somme des récompenses

- Interaction dans les DEC-POMDP

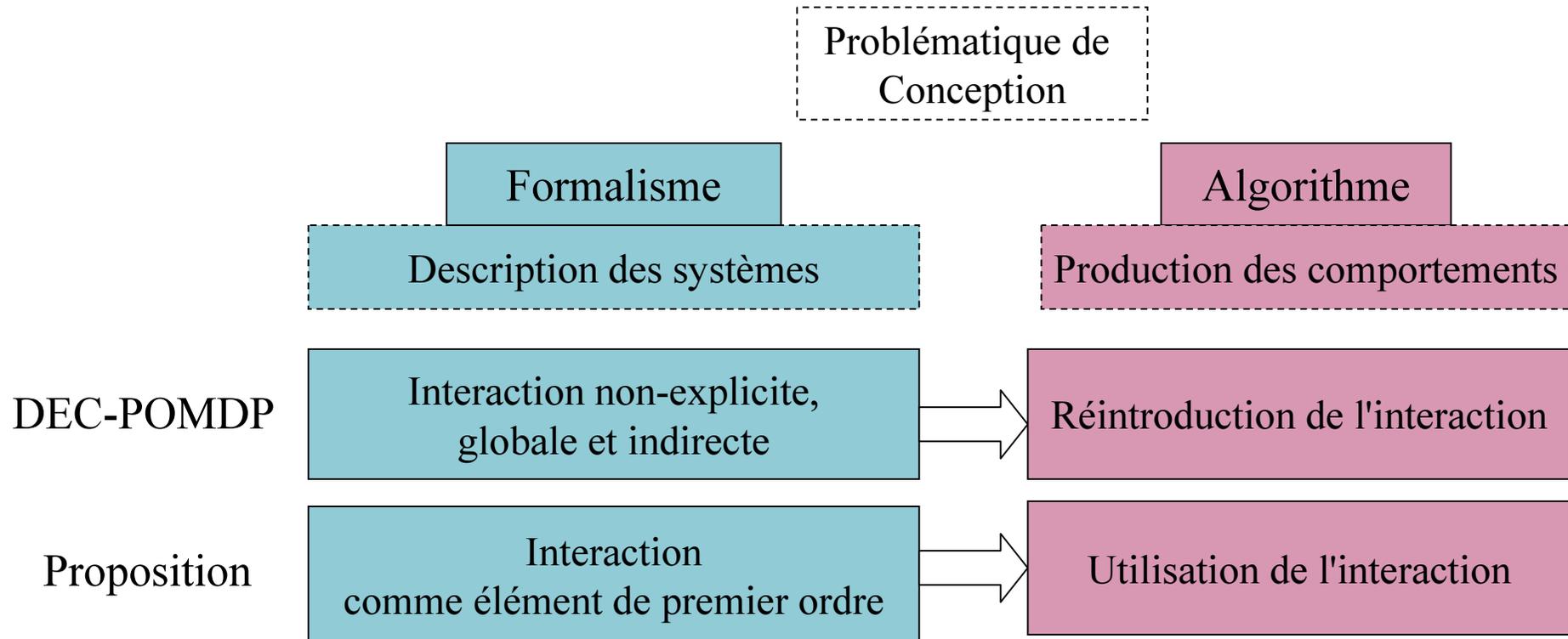
- Interaction indirecte → Difficile à évaluer localement
- Représentée globalement

# Résolution d'un DEC-POMDP

- Deux approches
  - Centralisée : Construction globale
    - → Problème a été prouvé NEXP
  - Décentralisée : Construction locale
    - → Adaptation sans vue globale
- Problèmes liés aux approches décentralisées :
  - Co-évolution : comment apprendre quand l'autre évolue ?
  - *Credit assignment* : comment répartir les récompenses ?
- Plusieurs approches vues selon l'interaction
  - Guider l'apprentissage pour apprendre les interactions [Buffet03] [Chades02]
  - Structurer le système pour en tirer parti [Guestrin03] [Schneider99]

Introduire dans l'algorithmique l'interaction  
absente dans le formalisme

# Bilan des modèles markoviens



- Quelle interaction (réactive) ?
  - Paradoxe apparent
    - Comment construire des agents sociaux à partir d'agents réactifs sans représentation complexe ?

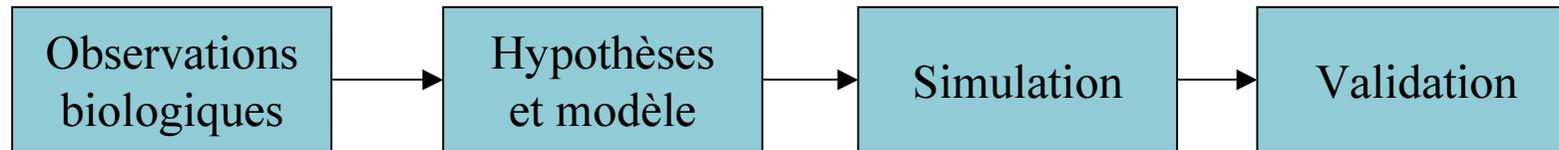
# Plan

- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

Quelles formes d'interactions permettent de tirer parti d'adaptations individuelles ?

# Démarche d'inspiration biologique

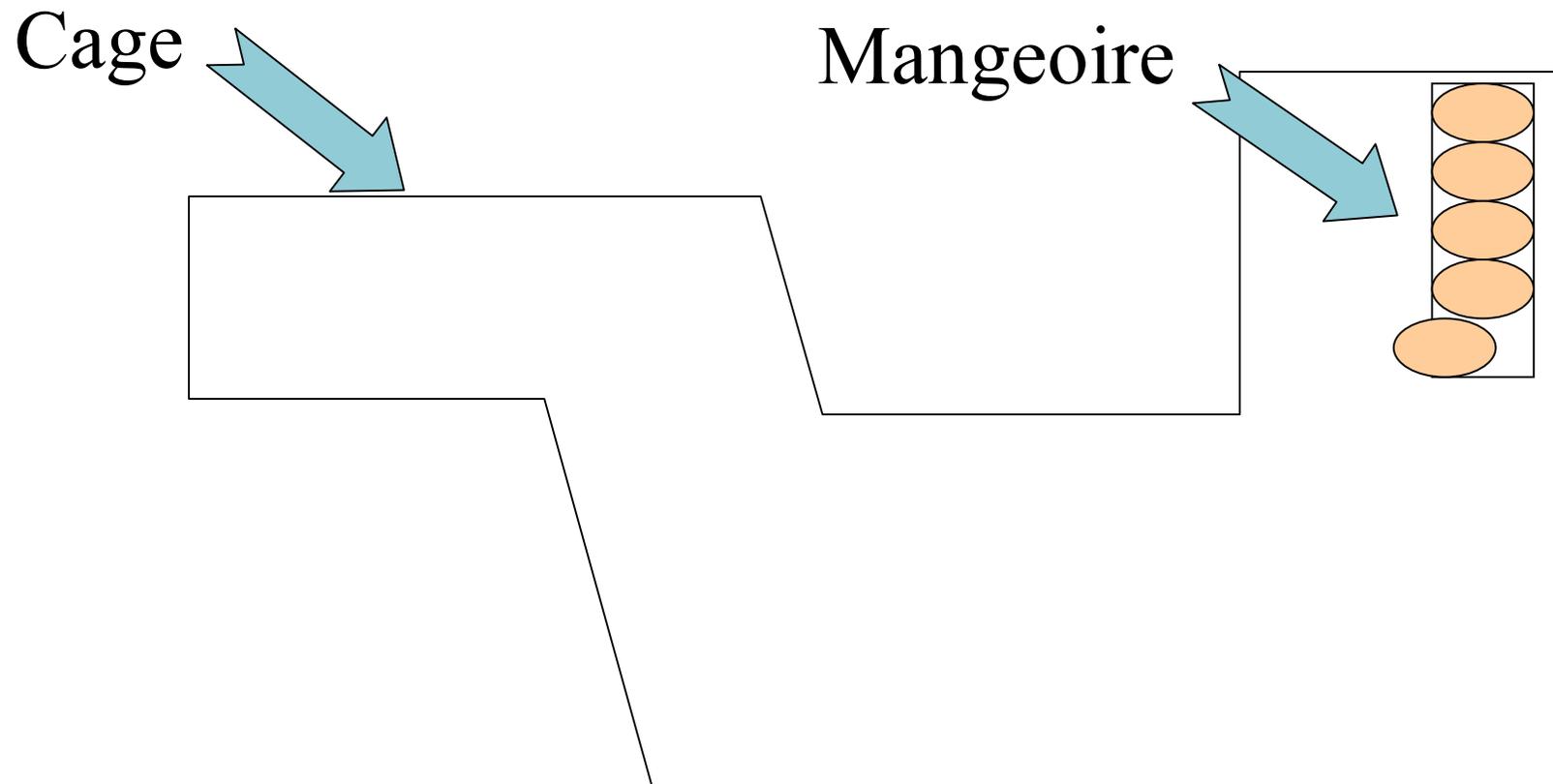
- Construction de modèles



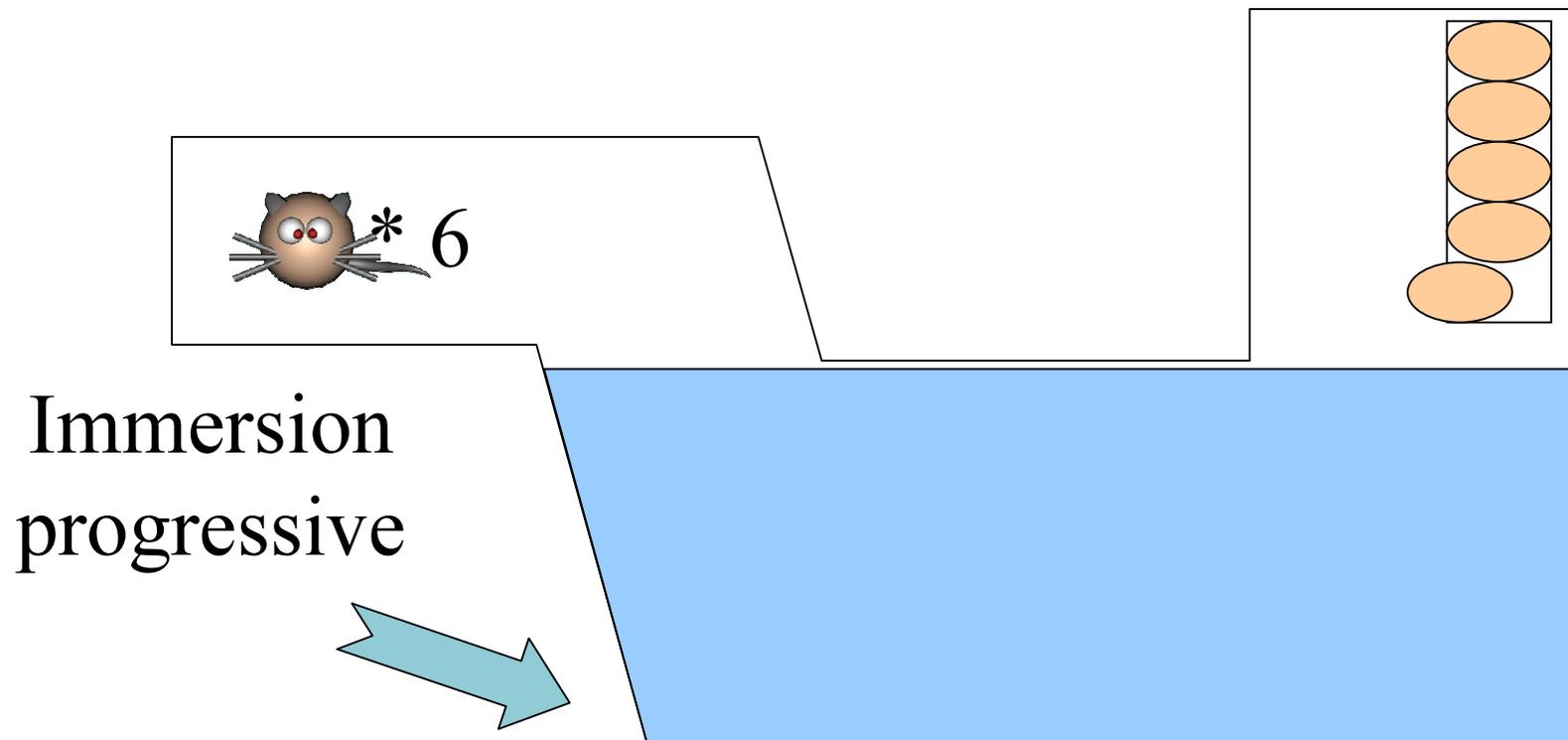
Métaphore constructive [Picault01]  
*"Construit un concept calculable"*

- Positionnement dans la thèse:
  - Inspiration biologique = démarche
  - L'objectif est de mettre en évidence un concept calculable :
    - une forme d'interaction utile
- Phénomène observé :
  - Trouver un phénomène collectif en éthologie
    - Les individus sont en interaction
    - Les individus savent répondre individuellement
    - Les individus exhibent des comportements différents en groupe
  - Spécialisation dans des groupes de rats [Desor91]

# Observations Biologiques

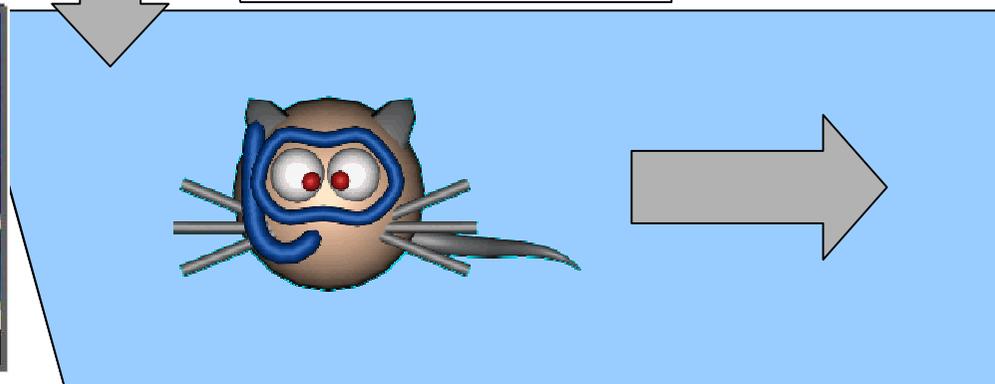
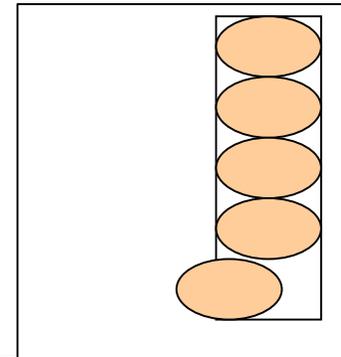
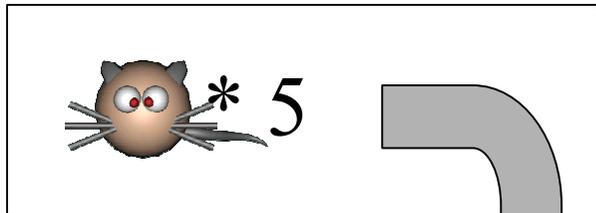
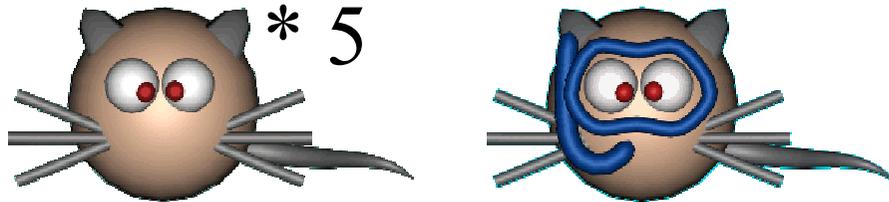


# Observations Biologiques

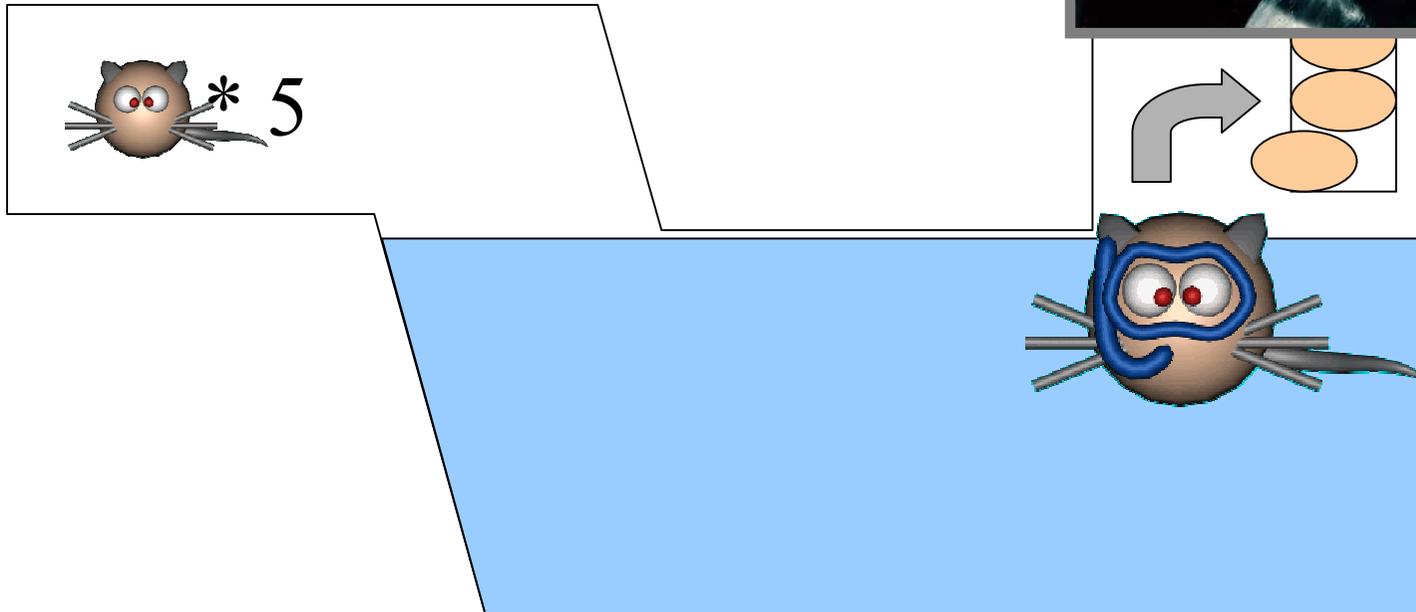
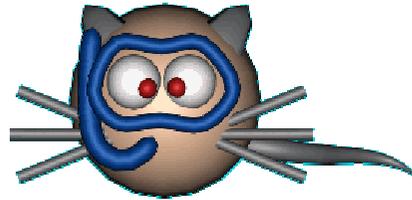
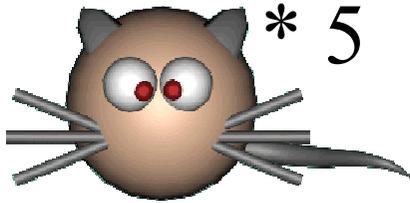


Immersion  
progressive

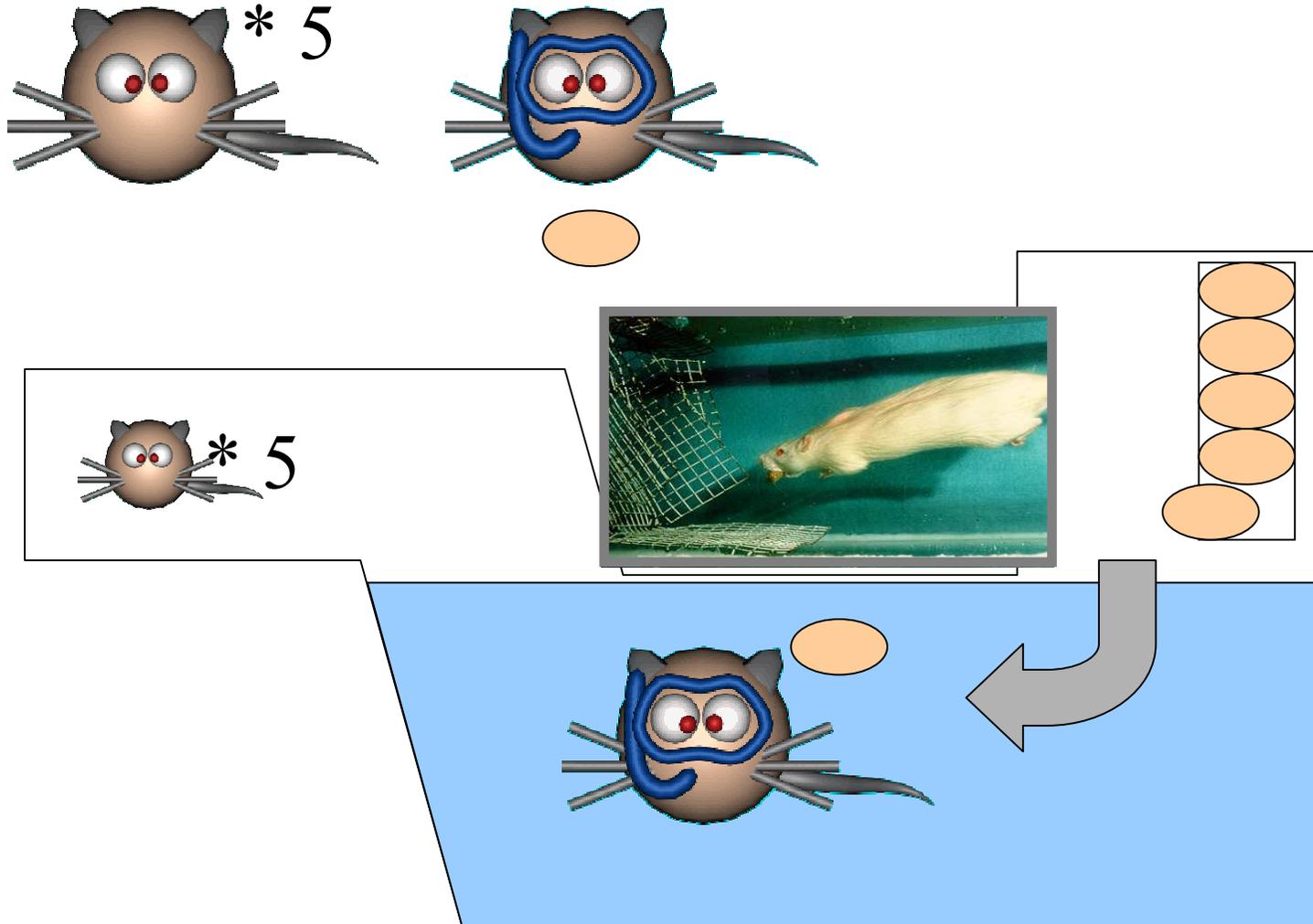
# Observations Biologiques



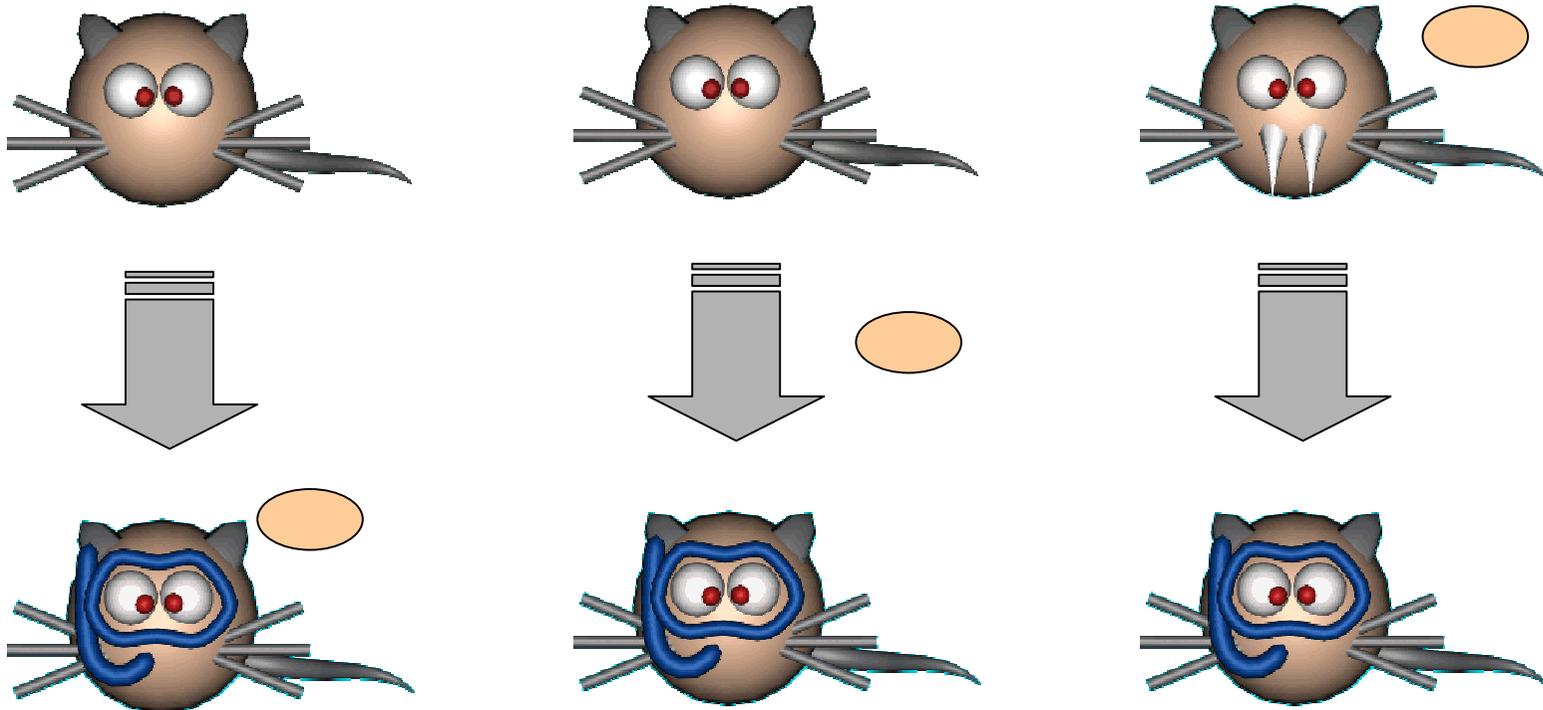
# Observations Biologiques



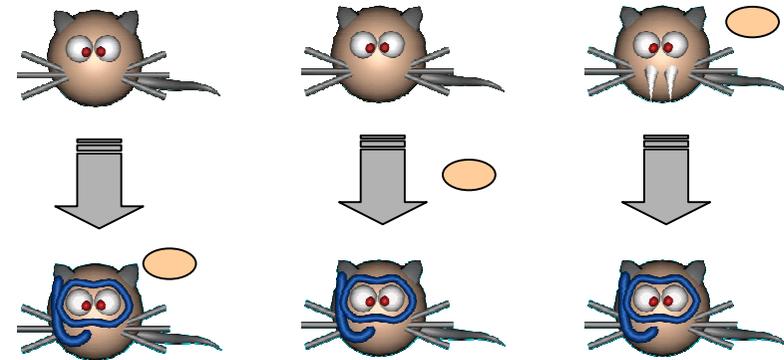
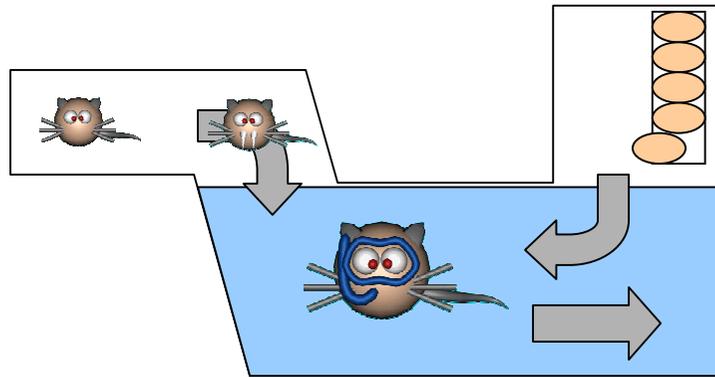
# Observations Biologiques



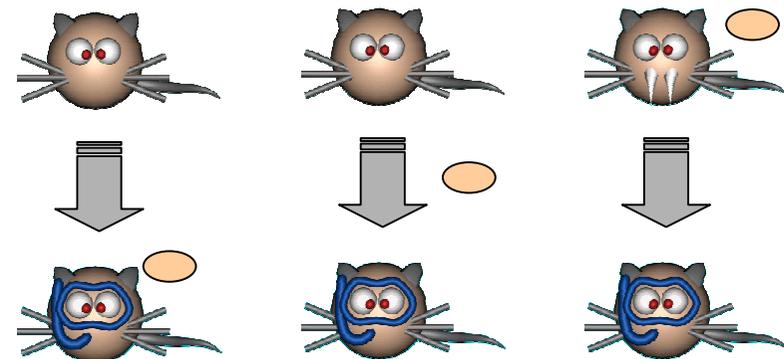
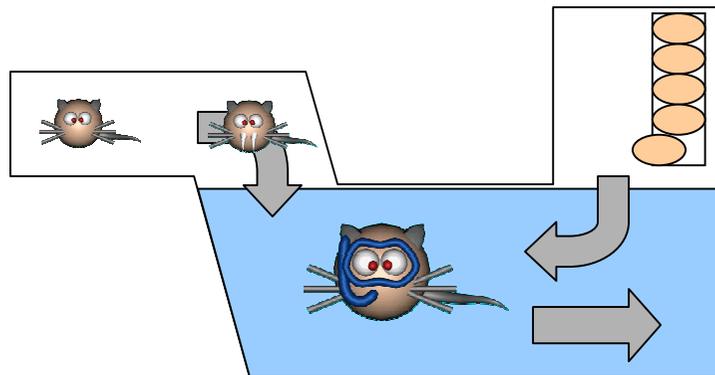
# Observations Biologiques



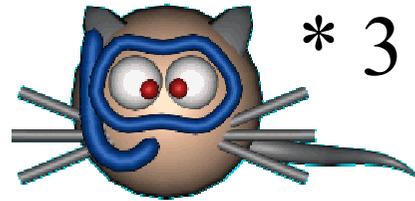
# Observations Biologiques



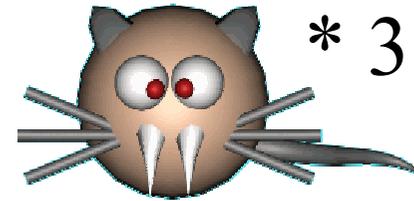
temps



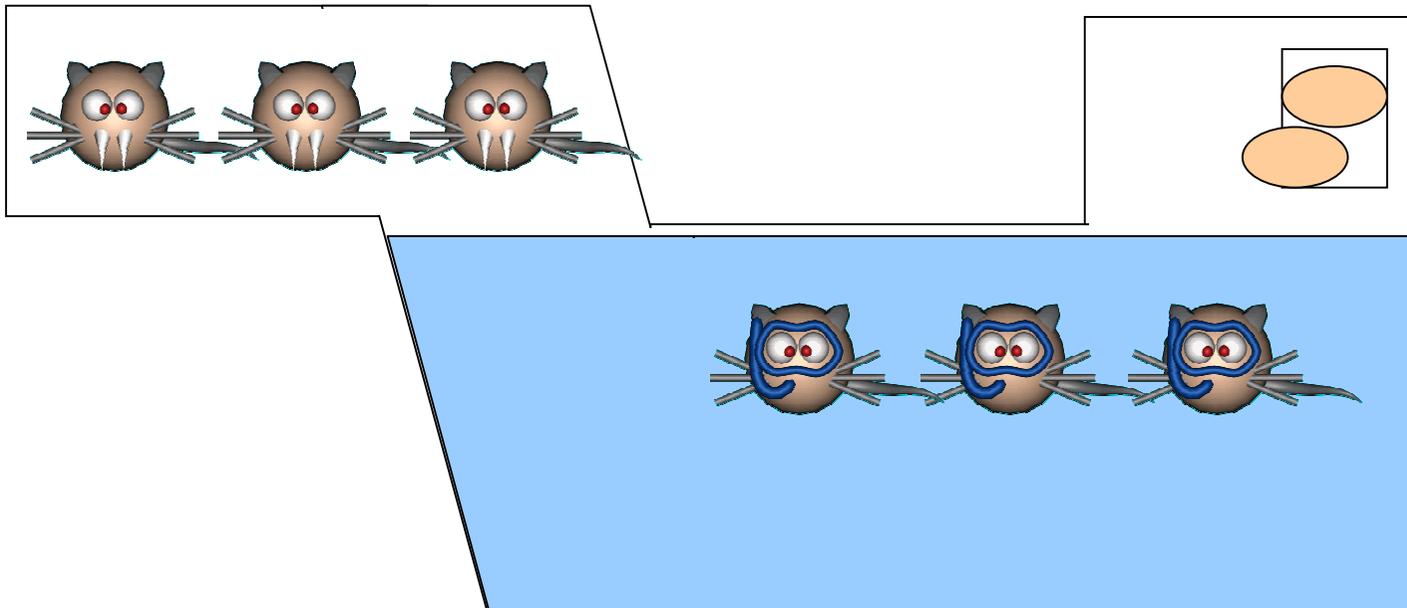
# Observations Biologiques



Rat  
transporteur



Rat  
non transporteur

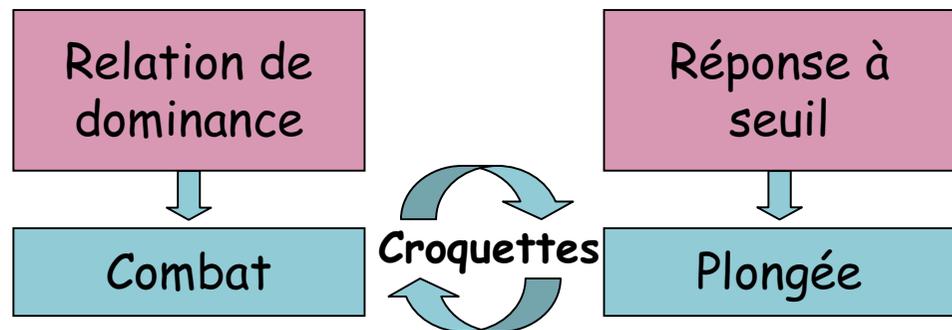


# Hypothèses

- Principe de parcimonie / rasoir d'Ockam
  - "Les explications les plus simples sont les meilleures"
  - Les individus n'ont pas de représentation complexe des autres
    - Interaction réactive
- Modèles de spécialisation existants
  - Réponses à seuil [Bonabeau99]
  - Relations de dominance [Hemelrijk99]

- Modèle original Hamelin
  - Coupler les deux

Une contribution de la thèse



# Modèle de simulation Hamelin

- Système Multi-agents

- Environnement

- Taille des croquettes

- Agents réactif : rat

- Anxiété  $\theta$
- Force  $F$
- Faim  $f$
- Croquette
- Pas de cognition sociale

Individuel

- Lois d'évolution

- Plongée

- Réponse à seuil [Bonabeau99]

$$T_{\theta}(f) = \frac{f^2}{f^2 + \theta^2}$$

$\theta$  Renforcée

- Manger

- Automatique

- Interaction

- Vol de croquette

Collectif

- Vol

- Relation de dominance [Hemelrijk99]

$$P_i(win) = \frac{F_i}{F_i + F_j}$$

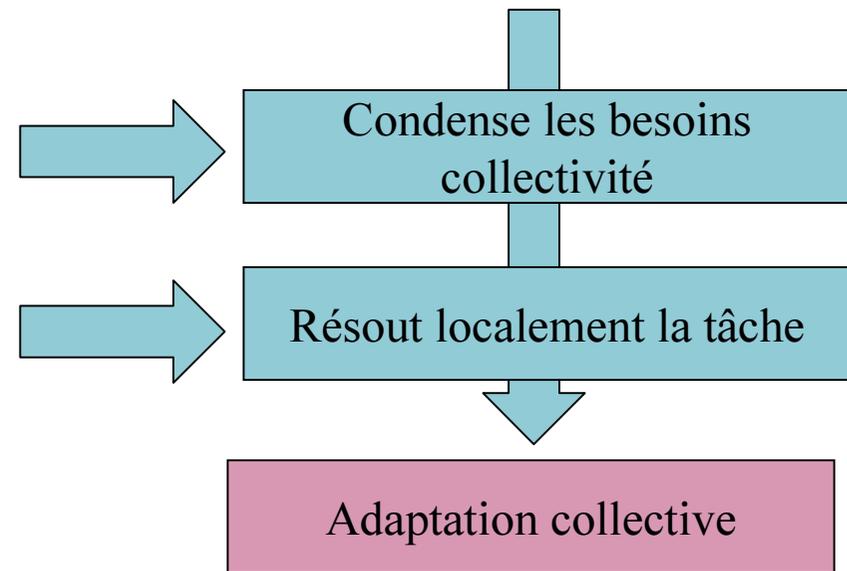
$F$  Renforcées

- Organisation

- Spécialisation

# Validation et interprétation

- Validation
  - Reproduction de la spécialisation
    - Mêmes profils
    - Mêmes proportions
  - Autres propriétés observées in vivo par les éthologues
- Interprétation
  - Interaction directe renforcée
    - Transfert de nourriture
  - Adaptation individuelle
    - Apport de croquette



# Bilan de l'inspiration biologique

- Objectif fixé:
  - Trouver une nouvelle forme d'interaction
- Réponse:
  - Interaction directe
    - Initiative individuelle
    - Résolution collective et évolution décentralisée
    - Sans représentation complexe des autres
  - Tirer parti de processus d'adaptation individuels
- Hamelin est une simulation et n'est que descriptif
  - Pas de tâche à résoudre
  - Comportements a priori

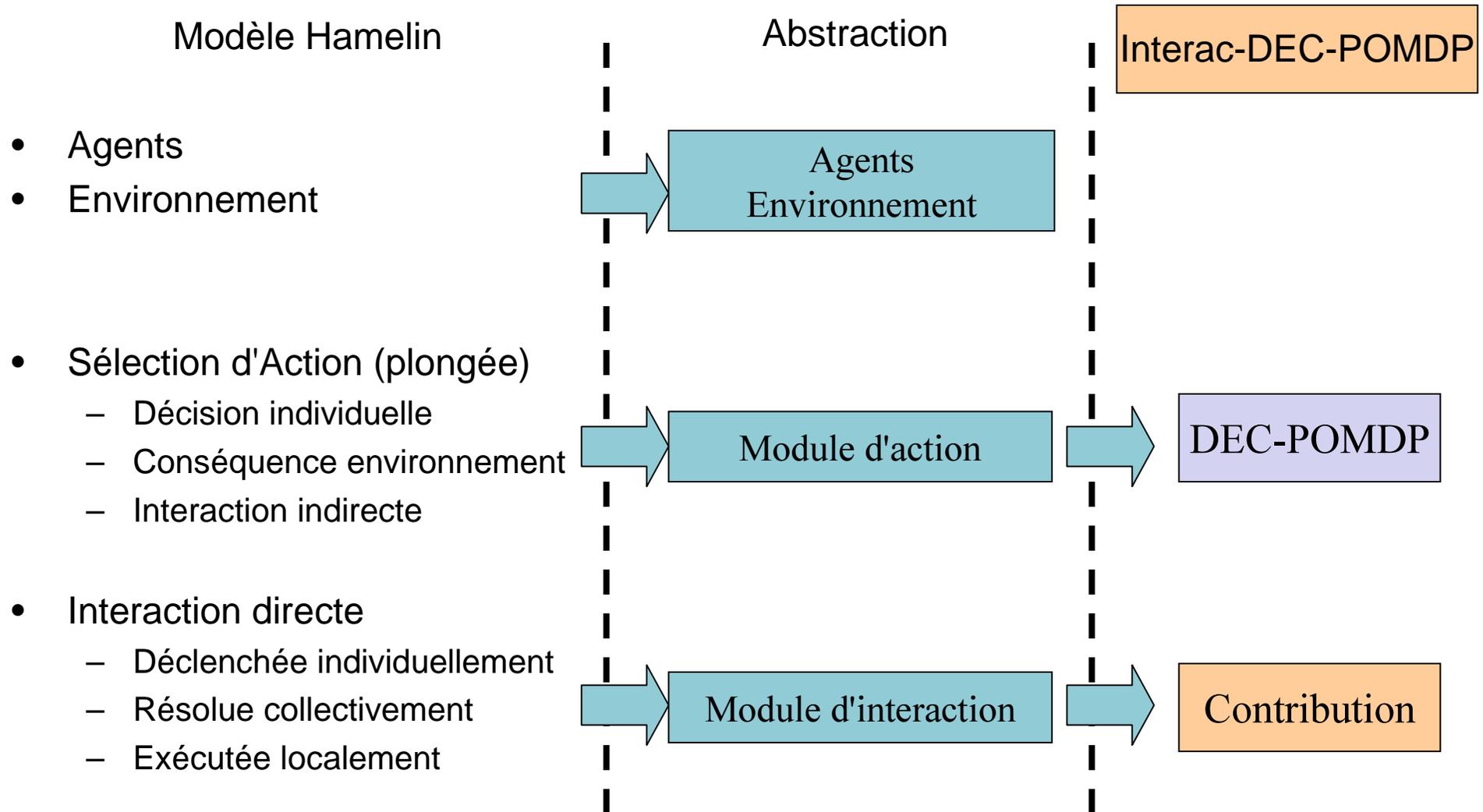
Comment faire de l'interaction directe un principe de résolution générique ?

# Plan

- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

Comment représenter  
ces interactions directes ?

# Abstraction de Hamelin

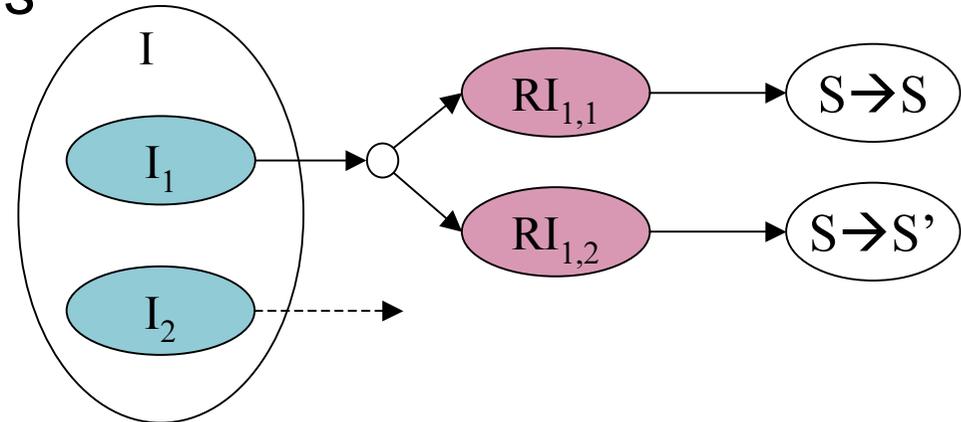


# Description des interactions

- Structure des interactions

- Type d'interactions
- Résultats d'interactions
- Transition de résultats

Notion de rôle



- Choix individuel
  - Type d'interaction
  - Partenaires interactions
- Résultat décidé collectivement

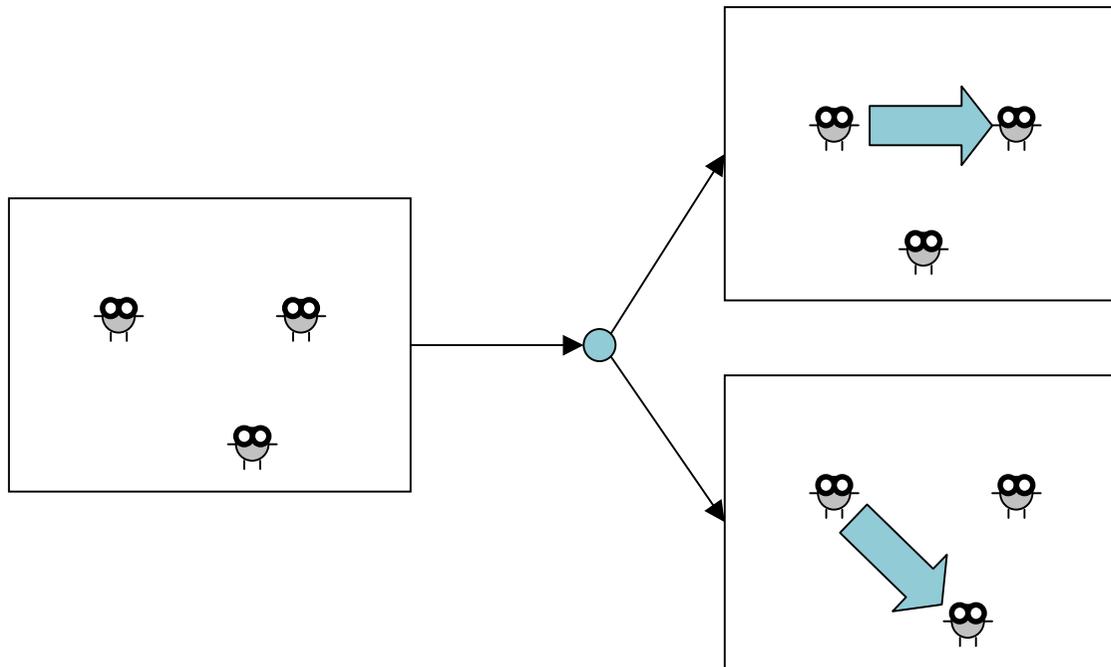


Capacité à reconfigurer l'interaction à l'exécution

# Exécution des interactions

- [Quand et avec qui interagir ?]
- Déclenchement
  - Décision individuelle
  - Choix des partenaires d'interaction
  - Choix de l'interaction

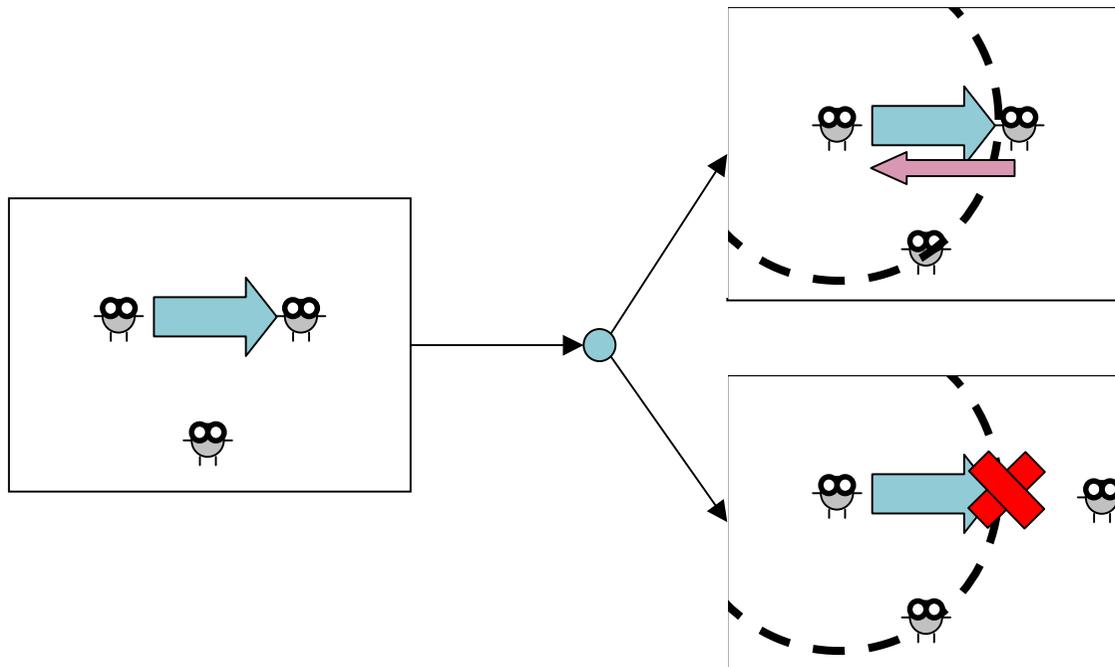
$$\pi_{decl,i} : \Gamma_i \rightarrow I \times agent^*$$



# Exécution des interactions

- [La communication est-elle possible ?]
- Fonction *possible*
  - Simule les capacités de communication
  - Rend compte si l'interaction est instanciable

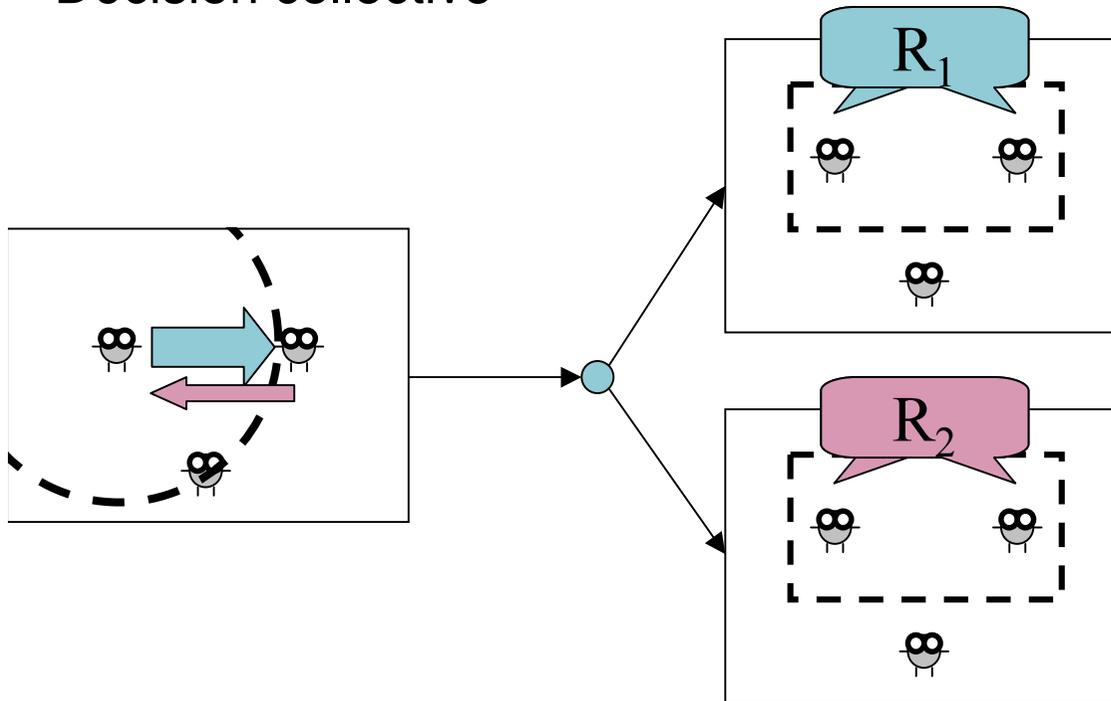
$$poss : S \times I \times agents^* \rightarrow \{0, 1\}$$



# Exécution des interactions

- [Quel Résultat choisir ?]
- Politique de résolution
  - Choix d'un résultat
  - Communications locales
  - Décision collective

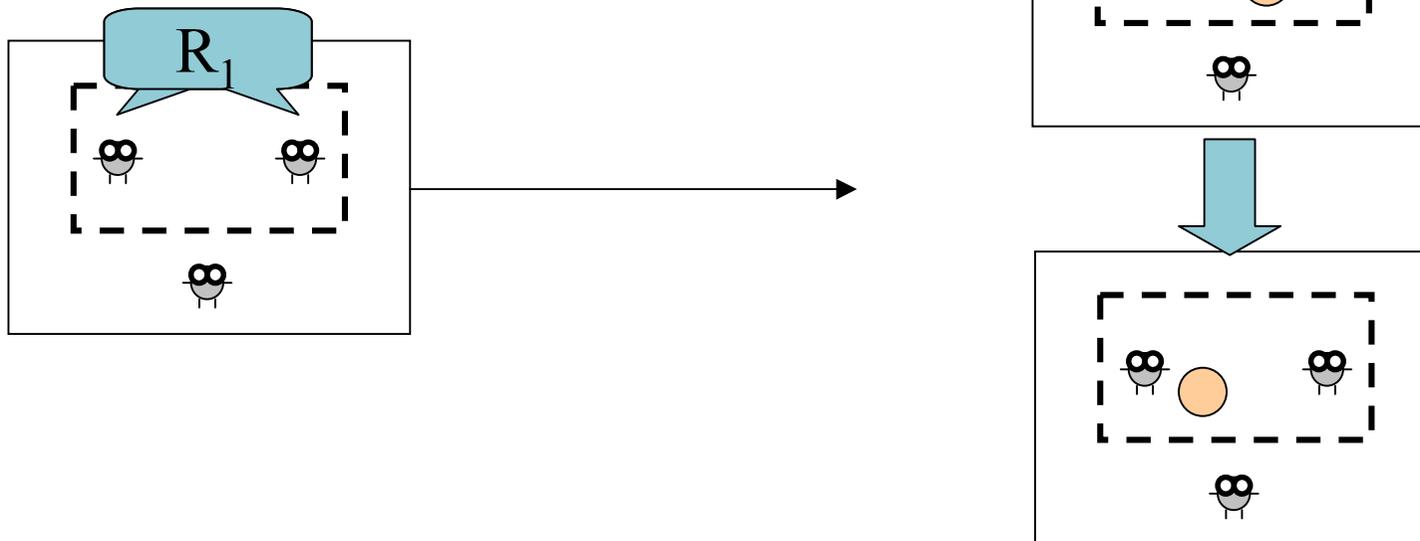
$$\Pi_{agent^*, I_k} : \Gamma_{agent}^* \rightarrow RI_k$$



# Exécution des interactions

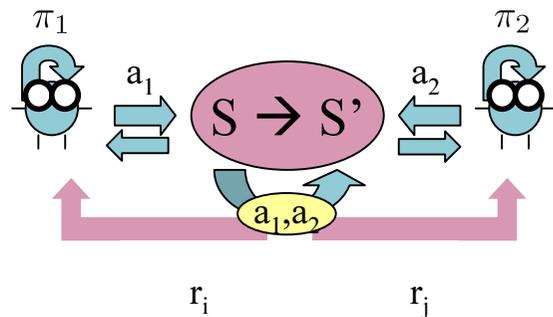
- [Quelle évolution ?]
- Exécution du résultat
  - Matrice de transition locale
  - Dépend des agents

$$TRI_{k,l} : agents^* \times S \rightarrow S$$



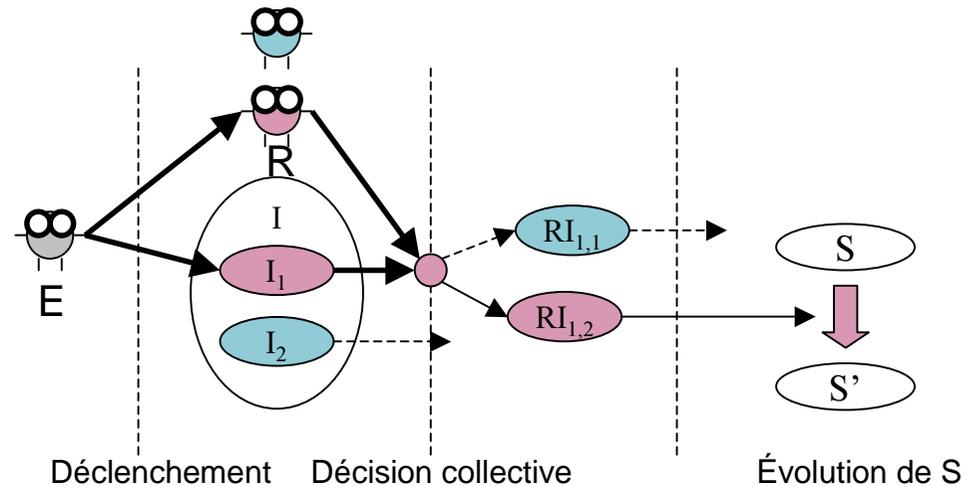
# Exécution Interac-DEC-POMDP

## Module Action



- Pour tout agent
  - Observation  $o_i$
  - Action  $a_i = \pi_i(o_i)$
- Exécution action jointe
  - $s' \leftarrow T(s, a)$
- Récompense
  - $r_i \leftarrow r_i(s, a, s')$

## Module Interaction



- Séquentiellement pour chaque agent  $i$ 
  - Observation  $o_i \leftarrow O_i(s)$
  - Déclenchement  $(I_k, \text{agent}_j) \leftarrow \pi_{\text{decl},i}(o_i)$
  - Si possible alors
    - Observation  $o_j$
    - Communication locale
    - Choix du résultat  $Ri_{k,l} \leftarrow \prod_{i,i}(o_i, o_j)$
    - Exécution du résultat  $s \leftarrow TRI_{k,l}(S, i, j)$

# Problème Interac-DEC-POMDP

- Étant donné un Interac-DEC-POMDP

- Actions
- Interactions
- Communications locales

$\langle S, A_i, T, R, \Gamma_i, O, I, RI, TRI, \text{poss} \rangle$

- Déterminer

- [quoi faire ?]
  - Politique d'action pour chaque agent
- [quand et avec qui interagir ?]
  - Politique de déclenchement pour chaque agent
- [comment interagir ?]
  - Politique de résolution pour chaque couple

$\pi_i$

$\pi_{decl,i}$

$\Pi_{agent^*, I_k}$

- Pour maximiser la somme des récompenses collectives
- Interac-DEC-POMDP contient les DEC-POMDP
  - Recherche optimale au moins aussi complexe
  - DEC-POMDP est NEXP

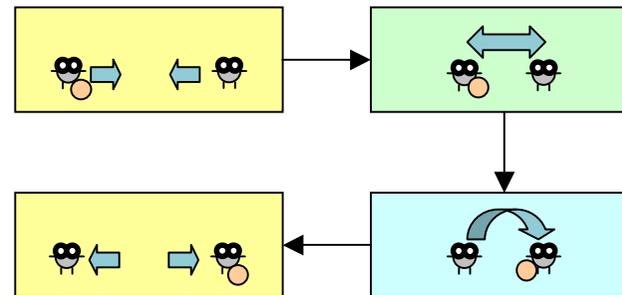
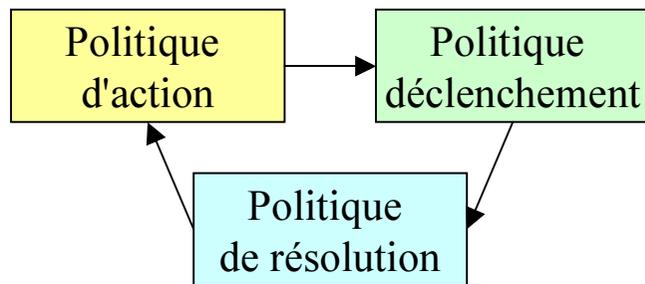
# Plan

- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

Comment utiliser  
ces interactions directes ?

# Proposition

- Objectif : calculer simultanément et localement
  - Quoi faire :  $\pi_i$
  - Quand interagir :  $\pi_{\text{decl},i}$
  - Comment interagir :  $\prod_{i,j}$
- Interdépendance des politiques

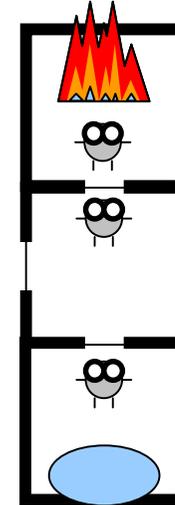


# Utilisation du formalisme

- Une sous-classe de Interac-DEC-POMDP

Générique
<ul style="list-style-type: none"><li>• Actions</li><li>• Espace global divisé</li><li>• Interaction locales</li><li>• Récompenses locales</li></ul>

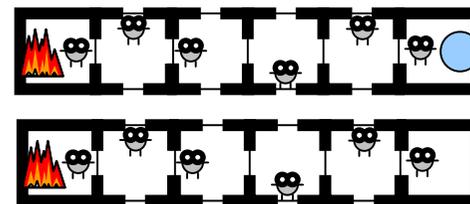
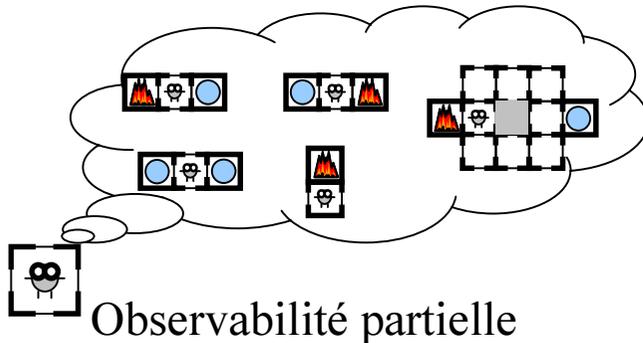
Instanciation : Pompiers
<ul style="list-style-type: none"><li>• <b>Déplacement</b></li><li>• Agent = limité à pièce</li><li>• Échange de seaux</li><li>• Seul agent près du feu perçoit R</li></ul>



- Sous problème
  - Centré sur interaction

Objectif :  
valider l'interaction directe

- Problèmes



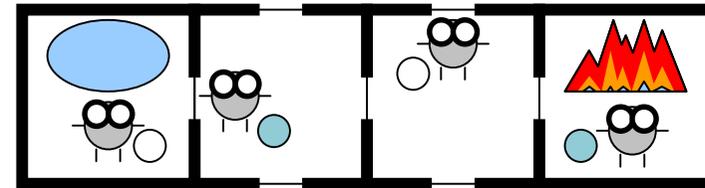
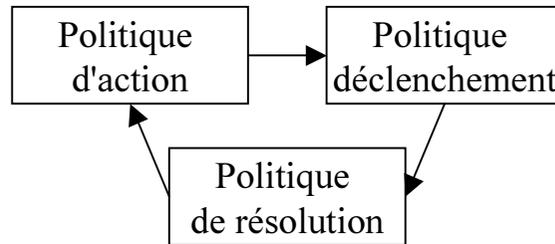
Problème purement collectif

# Représentation des politiques

- Apprentissage par renforcement dans un MDP
  - Q-learning [Watkins92]
    - Q-valeurs  $Q(s,a)$  = espérance de gain
      - partant  $s$ , effectuant  $a$ , suivant  $\pi$
    - Politique optimale paramétrée par Q-valeurs  $Q^*$ 
      - $\pi^*(s) = \operatorname{argmax}_a (Q^*(s,a))$
    - Apprentissage pour trouver  $Q^*$  suite à des expériences  $(s,a,s',r)$ 
      - $Q(s,a) = (1-\alpha) Q(s,a) + \alpha (r + \gamma \max_{a'} Q(s',a'))$
- Paramétrer les politiques d'un Interac-DEC-POMDP
  - $\pi_i$  par des  $Q_i(s_i, a_i)$
  - $\pi_{\text{decl},i}$  par des  $Q_{\text{decl},i}(s_i, I_k, \text{agent}^*)$
  - $\Pi_{i,j}$  par des  $Q_{\text{interac},i}(I_k, \text{agent}^*, s_{\text{agent}}, R_{k,l})$

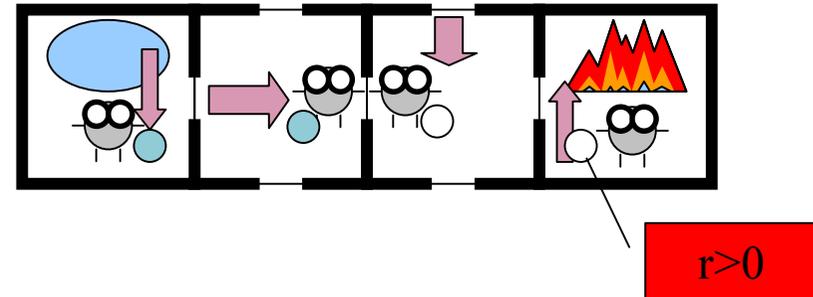
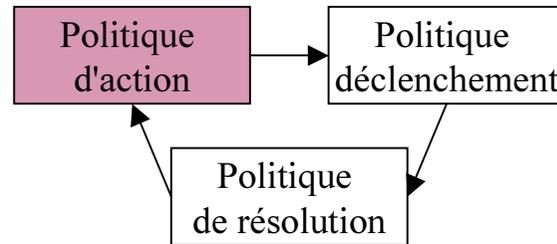
Individuelles

# Apprentissage des politiques



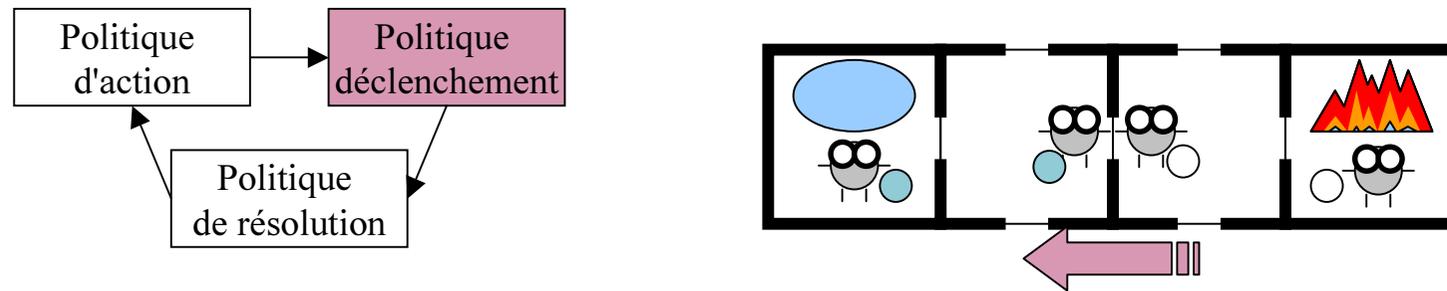
- État initial

# Apprentissage des politiques



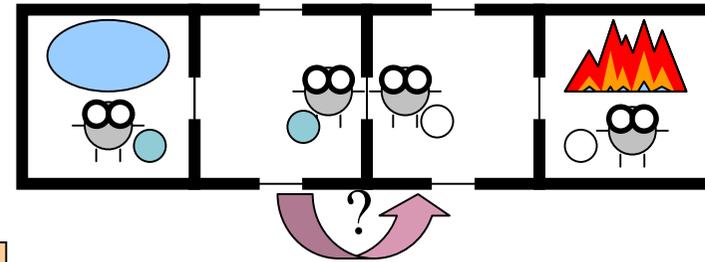
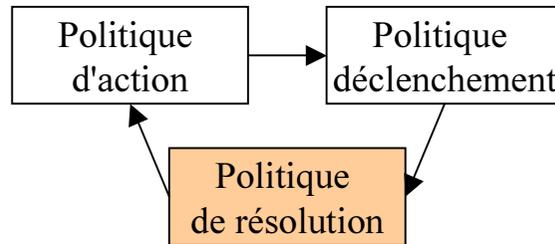
- Exécution des actions
  - A partir de politique d'action  $\leftarrow Q_i(s, a_i)$
- Récompenses individuelles

# Apprentissage des politiques



- Exécution des déclenchements
  - A partir de politiques de déclenchement  $\leftarrow Q_{\text{decl},i}(s,l)$
- Instanciation des interactions directes

# Apprentissage des politiques



- Résolution des interactions directes

- Objectif : Choisir un résultat

- A partir des politiques de résolution  $\leftarrow Q_{\text{interac},i}$

- Principe : heuristique inspirée de Hamelin

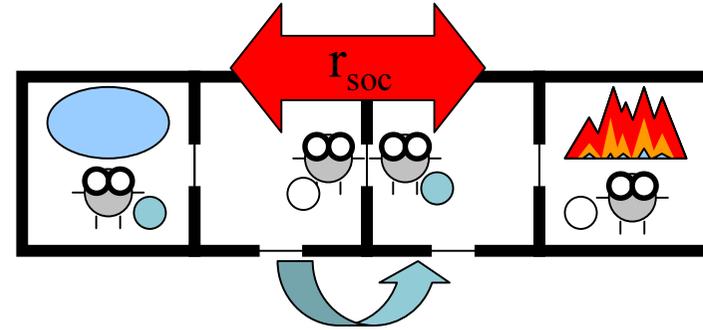
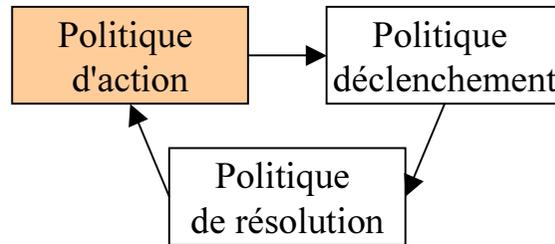
- Force  $\rightarrow$  Intérêt (évalué par Q-valeurs)
- Maximisation de la somme des intérêts des agents en interaction

altruisme

- Analyse

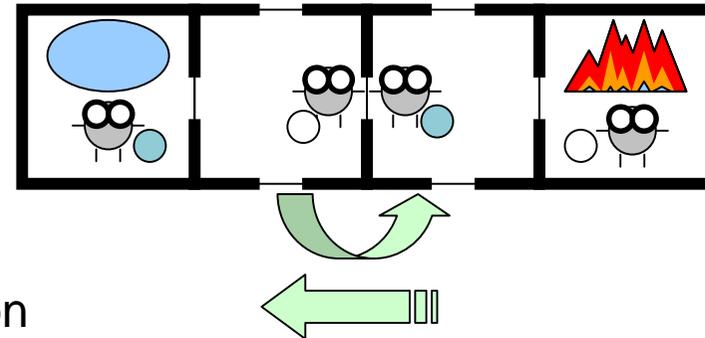
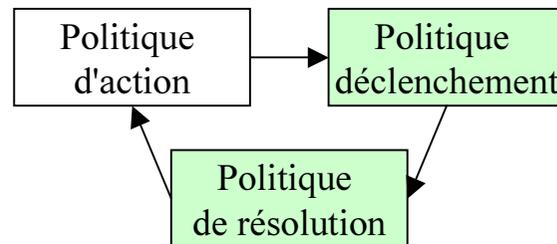
- Évaluation locale du gain pour le système  $\rightarrow$  Comportement collectif
- Lié à la tâche et aux actions futures par les Q-valeurs  $\rightarrow$  Générique
- Représente les politiques de manière distribuée  $\rightarrow$  décentralisé

# Apprentissage des politiques



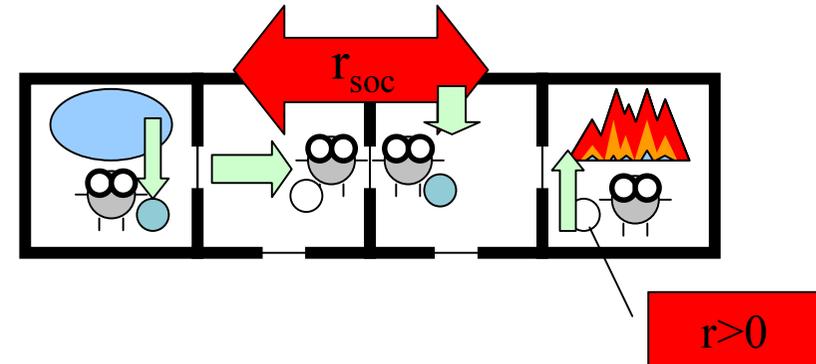
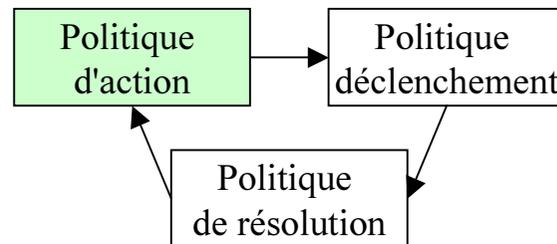
- Objectif :
  - Inciter un agent à reproduire une situation d'interaction
  - Même s'il ne reçoit pas directement de récompense
- Transfert de récompense
- Principe : Heuristique
  - L'interaction directe définit une nouvelle entité : les agents interagissant
  - Échange équitable de récompense au cours d'une interaction
- Interprétation
  - Partage automatique de tâche
  - Réponse au *credit assignment* : Les agents se guident mutuellement

# Apprentissage des politiques



- Mise à jour des politiques d'interaction
- Politique de résolution
  - A partir des états après interaction
- Politique de déclenchement
  - A partir du gain global dû à l'interaction directe exécutée
  - Évalué localement

# Apprentissage des politiques



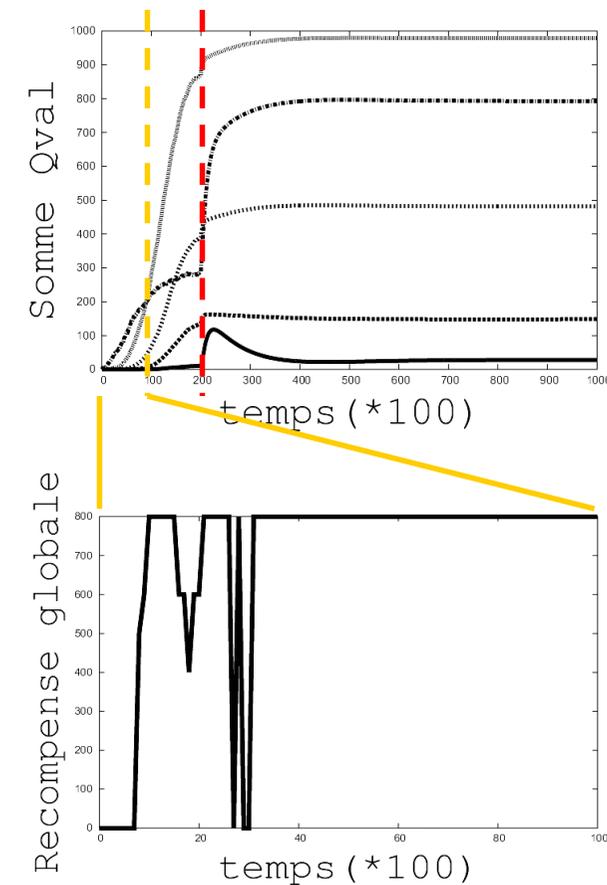
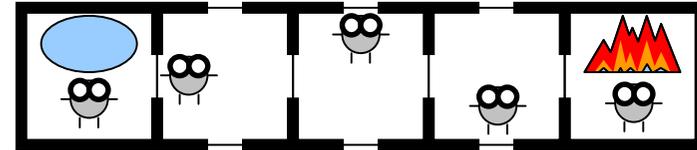
- Mise à jour des politiques d'action
- A partir de
  - État d'arrivé → Q-valeur d'arrivé
  - Action effectuée → Récompense individuelle
  - Interaction exécutée → Récompenses sociales

# Plan

- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

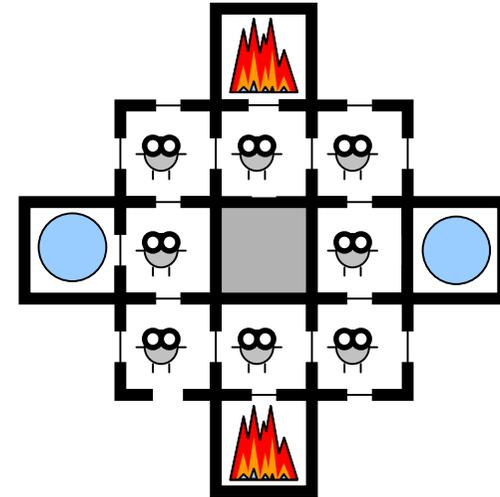
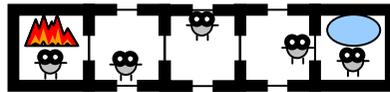
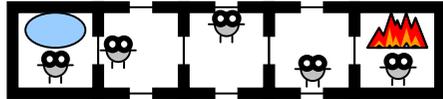
# Construction des comportements ?

- Validation algorithme
  - Cadre multi-agents
  - Transfers de récompenses
- Expérience simple
  - Exploration puis exploitation
- Résultats:
  - Apprentissage des politiques
    - action, déclenchement, Interactions
  - Convergence
    - Des comportements individuels
    - Du comportement global
  - Stabilité lors de l'exploitation



# Construction d'agents sociaux ?

- Plusieurs expériences

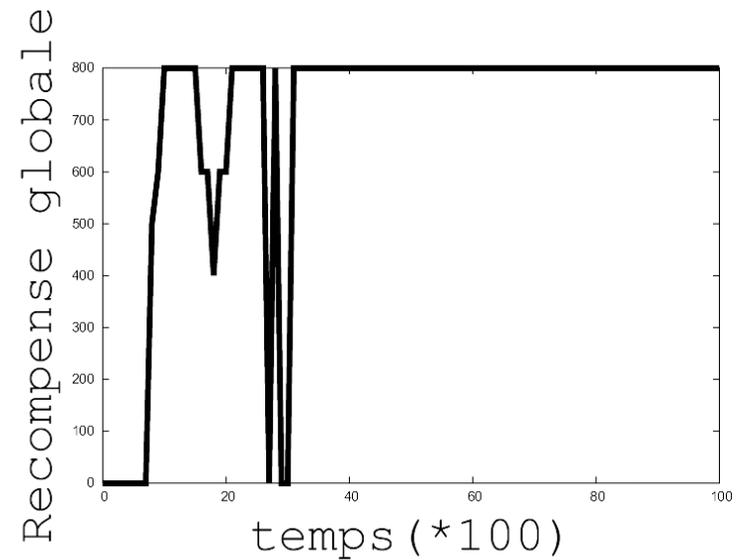
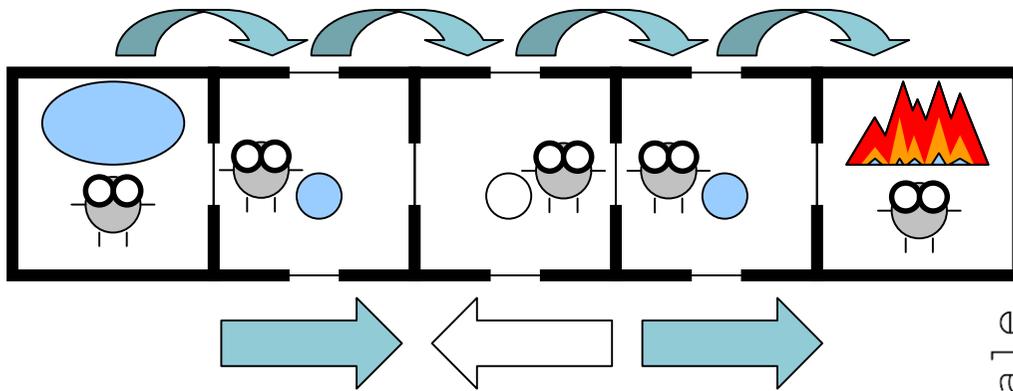


- Résultat : Comportements différents
  - Individuel :
    - Action différentes
    - Déclenchement différent
  - Collectif :
    - Résultats d'interaction différents
- Comportements individuels dépendent de l'environnement social :
  - Construit des agents sociaux

Capables de prendre en compte la présence  
d'autres agents sans les percevoir

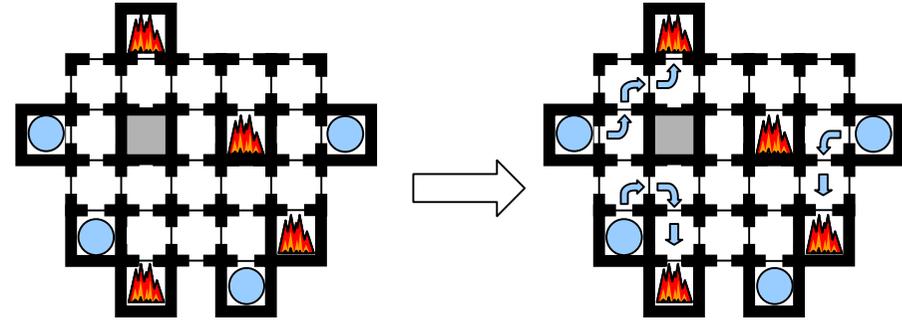
# Comportements collectif ?

- Pertinence :
  - Apparition d'organisations

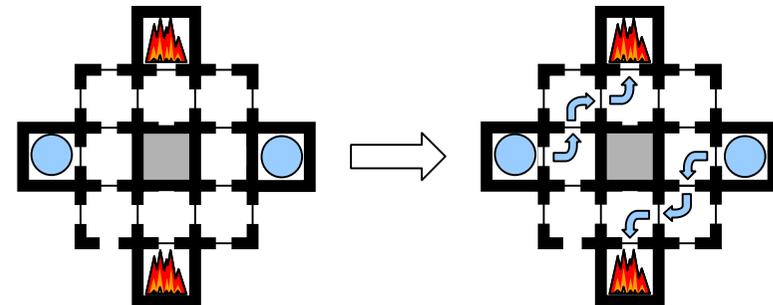


# Adaptation au nombre

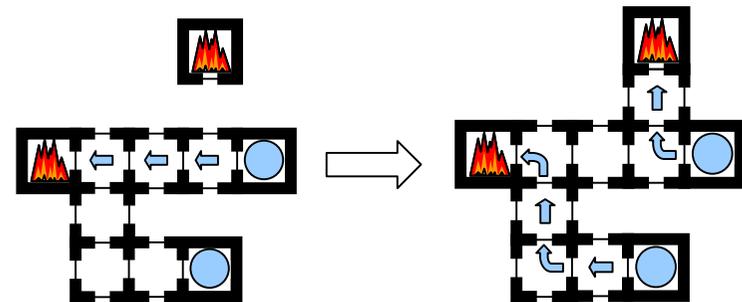
- Passage à l'échelle
  - Sous-optimal



- Brisure symétrie
  - Adaptation des organisations



- Réadaptation
  - Modifiant exploration

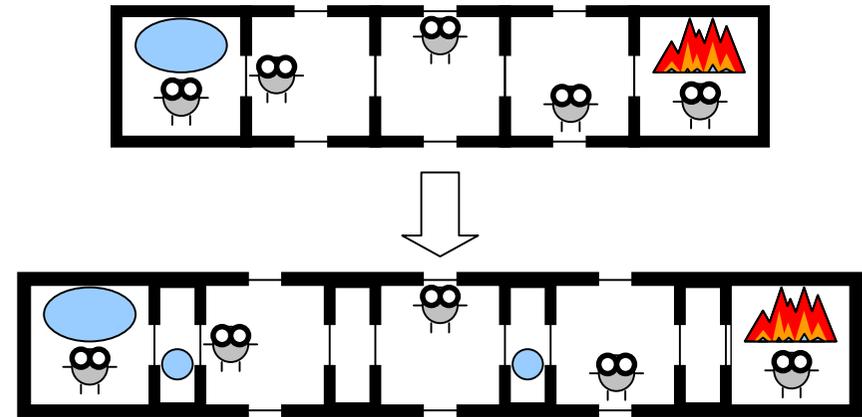


# Q-learning dans les DEC-POMDPs

- Comparaison avec le Q-learning utilisé sur des DEC-POMDPs

- Principe

- État : zones d'échanges
- Action : poser et prendre seau
- Perceptions partielles
- Récompenses globales



- Résultats :

- Problème du "*credit assignment*"

Intérêt de l'introduction de l'interaction :  
guider l'apprentissage des agents à l'exécution

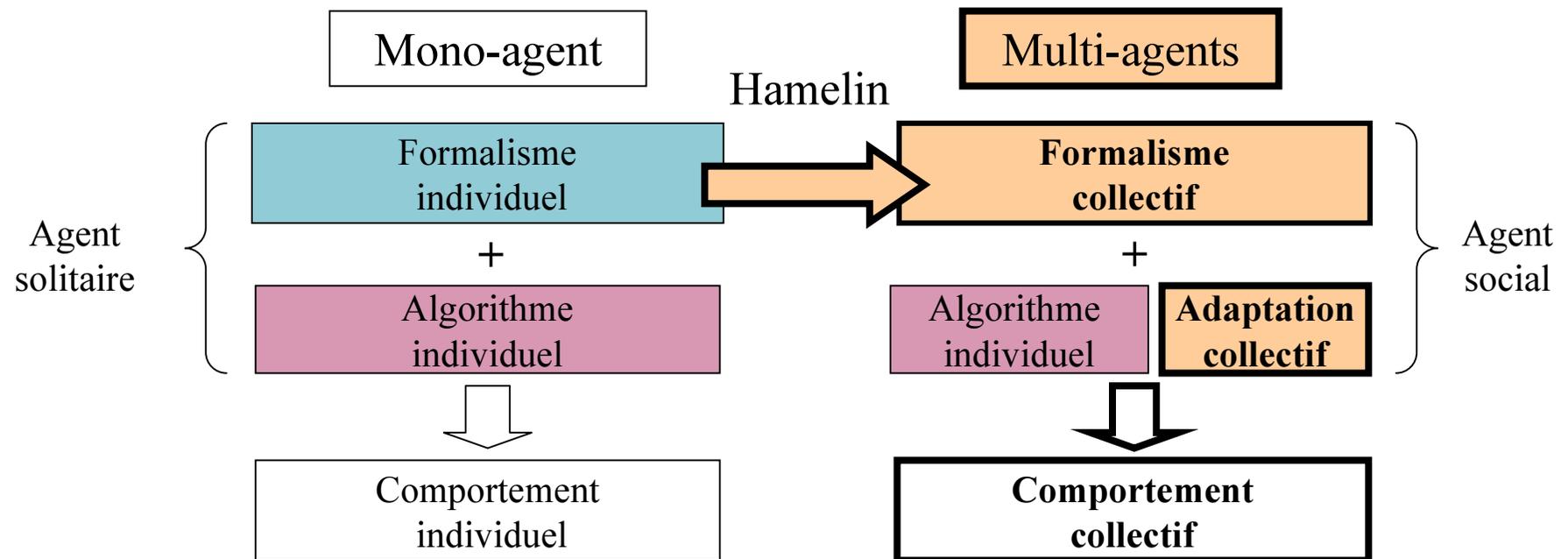
# Bilan

- Validation
  - Construction de comportements
  - Construction d'agents sociaux
  - Construction de comportements collectifs
  
- Limites
  - Heuristique
    - Sous-optimaux
  - Dilemme exploration/exploitation

# Plan

- Mise en regard mono-agent, multi-agents
  - Concepts
- Formalismes Multi-agents existants
  - Modèles markoviens
- Mise en évidence d'une nouvelle forme d'interaction
  - Inspiration biologique
- Notre proposition
  - Formalisme Interac-DEC-POMDP
  - Algorithme
- Validation
- Conclusion

# Démarche



# Contributions

- Un modèle original Hamelin
  - Simulation
  - Concept calculable d'interaction directe
- Un formalisme original Interac-DEC-POMDP
  - Générique pour représenter des systèmes multi-agents
    - Actions
    - Interactions directes reconfigurables
  - Définit une classe de problèmes multi-agents
- Un algorithme permettant de construire des agents sociaux
  - Construit automatiquement les politiques d'utilisation des interactions
  - Produit des organisations s'adaptant à la tâche

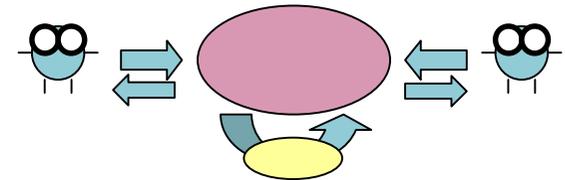
# Conclusion générale

- Formalisme : Alternative aux modèles actuels
  - Pour formaliser des problèmes SMA
  - Avec un concept fondamental dans ce domaine : l'interaction
    - Lien avec les méthodologies de conception
- Algorithme : Construction de comportements collectifs
  - De manière décentralisée
  - Avec des contraintes de localités fortes
  - Pour résoudre collectivement des problèmes
- Rationalité collective entre rationalité et interaction
  - Les agents ont la possibilité d'interagir
  - Les agents sont guidés par des fonctions de performance
- Apprentissage collectif décentralisé

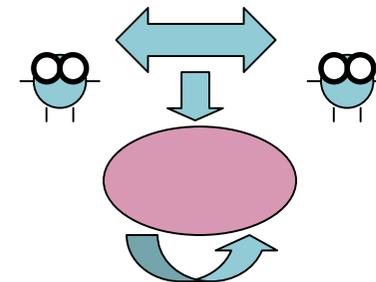
# Perspectives: De l'action aux structures

- Existant
  - Formalisme DEC-POMDP fondé sur l'action

- Apport de nos travaux
  - Formalisation explicite de l'interaction
    - → meilleures performances



- Questions :
  - Détecter les interactions utiles ?
  - Ajouter les interactions utiles incrémentalement ?



- A court terme
  - Meta-interactions

# Perspective: Des structures à l'organisation

- Existant
  - Approches OCMAS [Ferber02]
    - "Organisation Centered Multi-Agent Systems"
  - Méthodologies → organisations élémentaires
- Apport de nos travaux
  - Représentation et utilisation des structures collectives
- Question:
  - Operationalisation des approches OCMAS pour construire des organisations utiles ?
- A court terme
  - Interactions impliquant plus d'agents
  - Interactions dans le temps

