



Motivation intrinsèque



Motivation intrinsèque



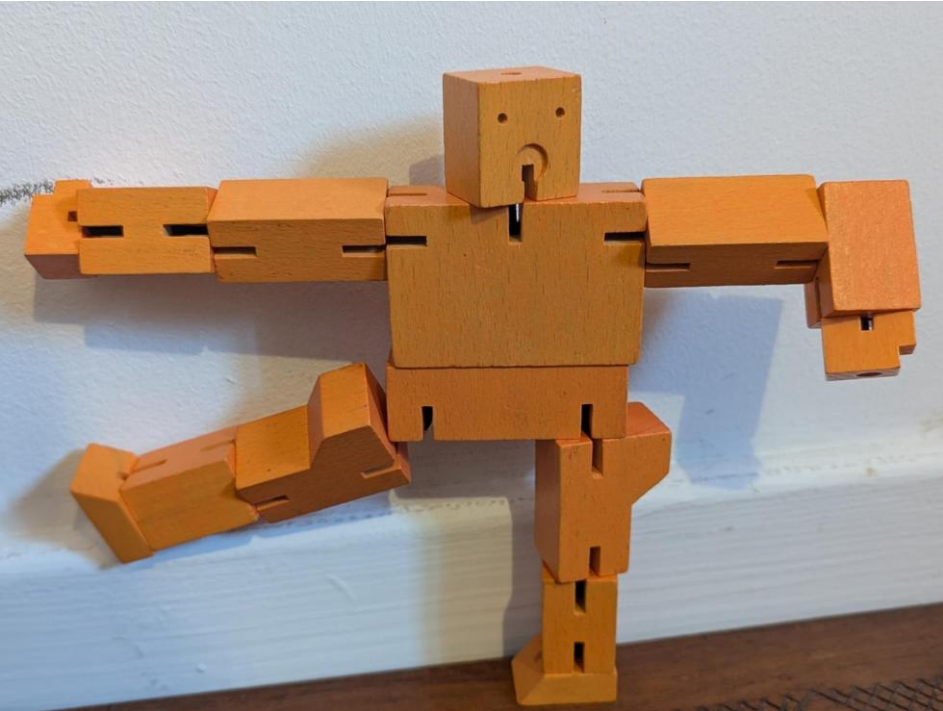
Pourquoi se lever le matin ?

“si tout est égal alors pourquoi faire
quelque chose plutôt que rien “



**en guise
d'introduction**

En guise d'introduction, trouvez les différences



plan

- 1) Histoire de la notion
- 2) Qu'est ce que la motivation intrinsèque ?
- 3) Comment est ce modélisé aujourd'hui ?
- 4) Zoom sur 2 recherches actuelles



Pourquoi ça bouge ?



Anomalie

Harlow (1950) : des singes résolvent des casse-têtes sans récompense.

White (1959) et Berlyne (1960), la curiosité.



Outil

Cybernétique (Wiener) et théorie de l'information (Shannon), 1948 :
Comparaison entre deux états



Psychologie

Deci (1971), les récompenses
érodent l'intérêt :
théorie de l'autodétermination.

Rescorla-Wagner (1972)

Sutton et Barto

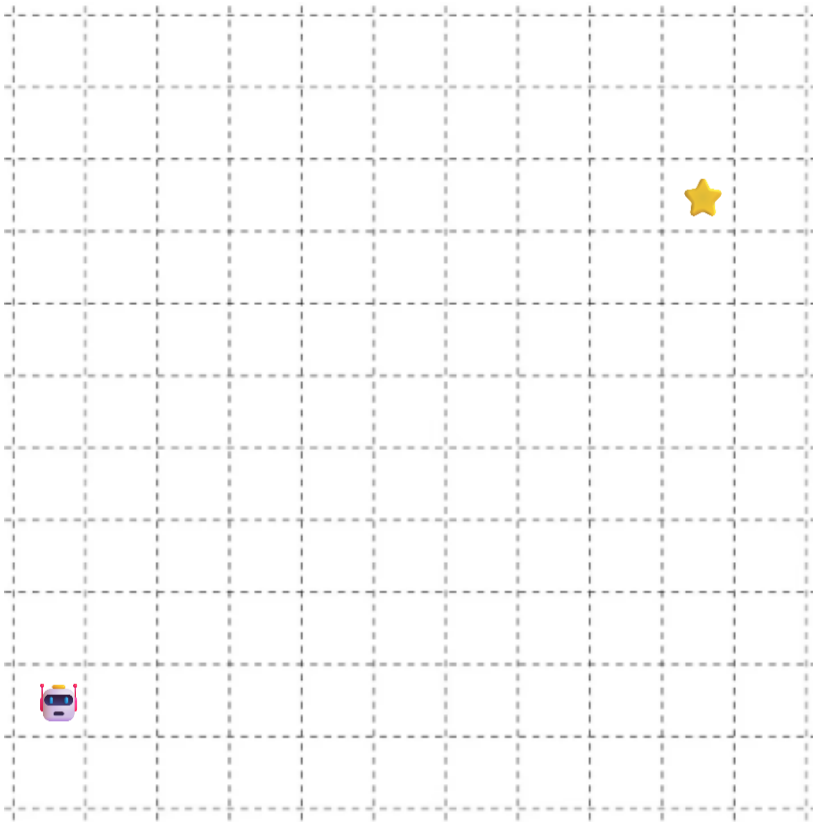
Schultz, Dayan et Montague (1997)



IA Symbolique

le but est donné de
l'extérieur.

Le piège de la récompense externe



Années 2000, retour vers la psychologie



Schmidhuber (1991)

Barto, Singh et Chentanez (2004)

Oudeyer et Kaplan (2007)

La motivation intrinsèque est la tendance à agir pour la satisfaction inhérente à l'activité elle-même, plutôt que pour obtenir une récompense venant de l'extérieur.

Concept psychologique

Fonction de récompense interne

Années 2010, Deep learning et boucle

IA

agents Atari (2013-2015),
bonus de curiosité (Bellemare
2016,
Pathak 2017, Burda 2018).

Sciences du vivant

Kidd et al. (2012),
les bébés regardent ce qui n'est ni
trop prévisible ni trop surprenant.

3 articles, 3 étapes

Aubret et al.(2023)

Structurer

Quel bilan pour la motivation
intrinsèque aujourd'hui ?

Oudeyer et al.(2007)

Formaliser

Comment traduire la curiosité en
algorithme ?

Zheng et al.(2024)

Corriger

Comment résoudre les
limites de ces
méthodes

Vers la motivation en IA ?

Au commencement était un agent dans un environnement



Historiquement, un objectif clair est donné à l'agent.

L'agent a pour but d'atteindre un état objectif (ou maximiser une fonction de valeur) et doit trouver la stratégie pour y parvenir.

la motivation est fermée : elle contraint entièrement les décisions de l'agent et lorsque l'agent a atteint l'état voulu, il s'arrête.

Exemple : jouer aux échecs

Problématiques

Pour la construction d'agent,

Comment faire pour que les agents explorent **d'autres comportements** et acquièrent d'autres compétences lorsque l'objectif est **lointain ou difficile** “**sparse reward**” ?

pour l'I.A développementale,

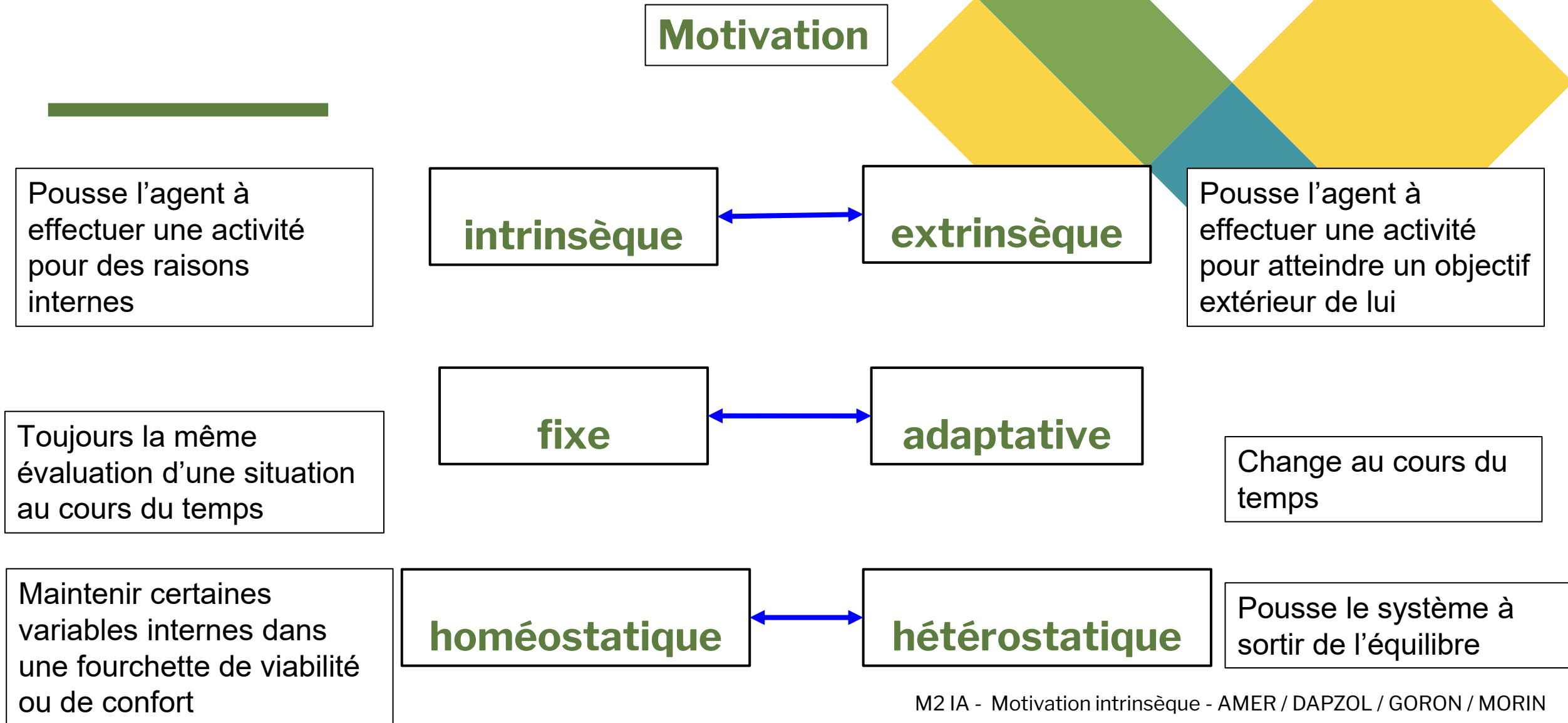
Comment faire pour que les agents fassent quelque chose plutôt que rien sans leur donner un objectif préétabli ?

Développer la créativité /
d'exploration

Permettre le développement de
compétences réutilisables

Qu'est ce que la motivation ?

→ différents axes d'analyse



Qu'est ce que la motivation ?

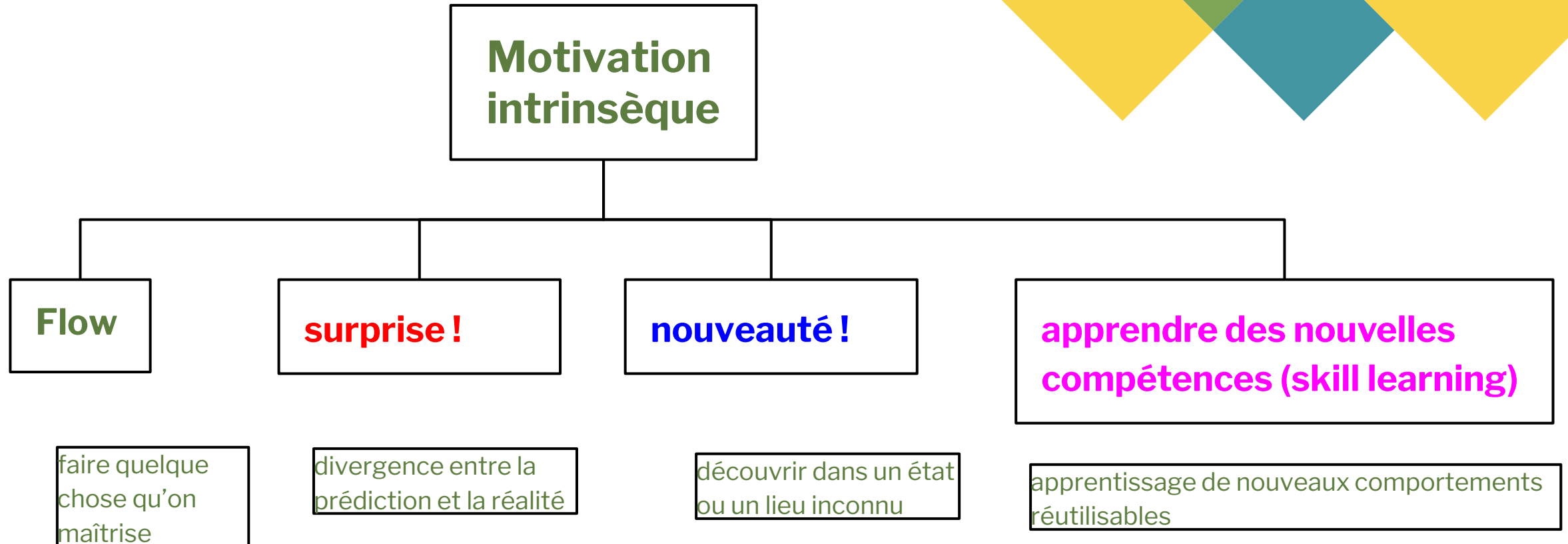
→ différents axes d'analyse

Pousse l'agent à
effectuer une activité
pour des raisons
internes

**Motivation
intrinsèque**



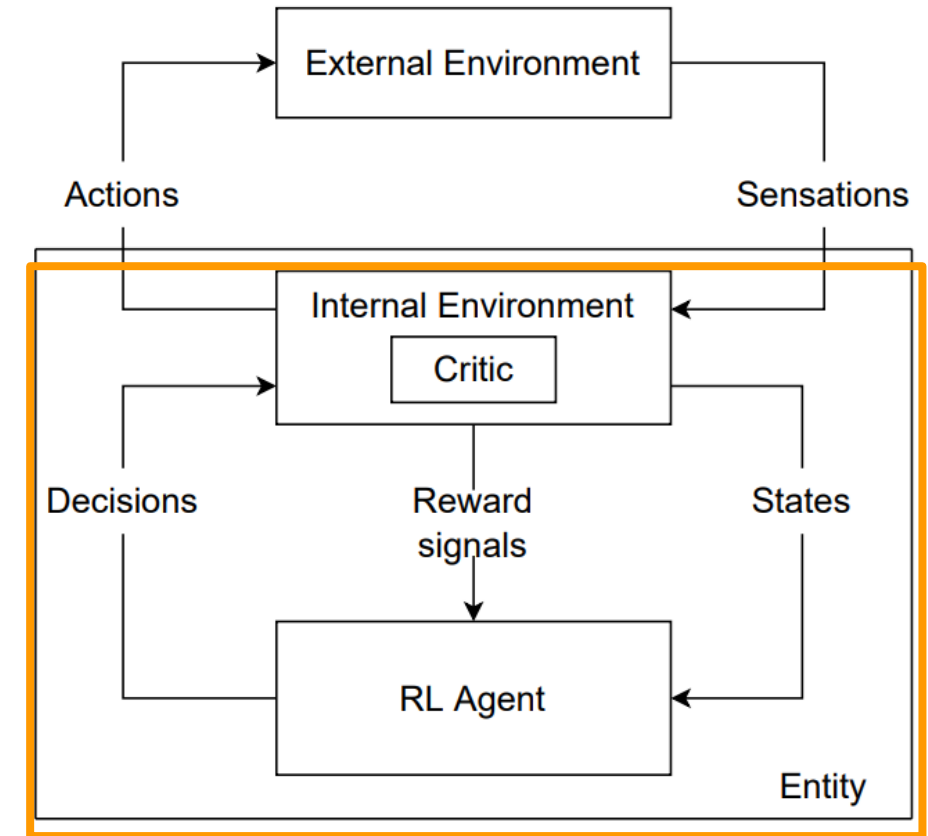
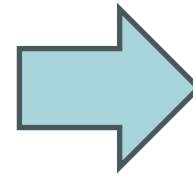
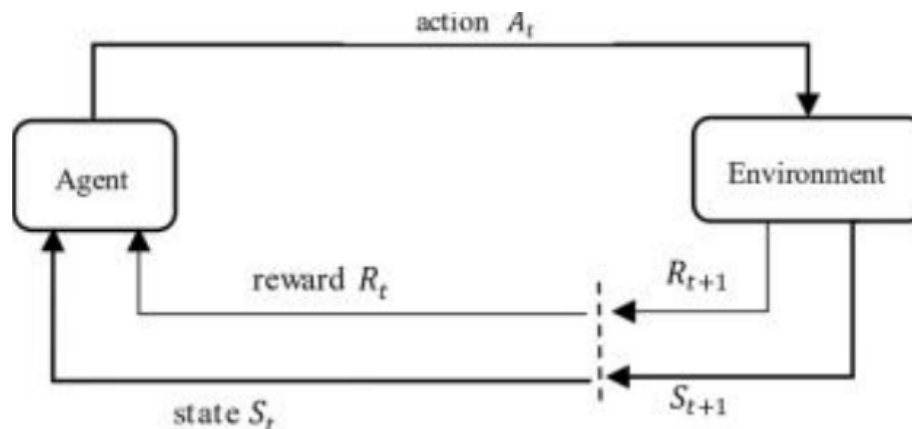
Qu'est ce que la motivation intrinsèque ?



Programme

- Introduction
- Qu'est ce que la motivation ? qu'est ce que la motivation intrinsèque ?
- **Quelles solutions pour la modéliser ?**
- Où est ce que c'est utilisé ? quels résultats actuels ?
- Quels défis aujourd'hui ?

- Mais comment modéliser ceci ?



En modélisant des processus et
des récompenses **internes**

Mais comment modéliser ceci ?

La surprise



“ je ne m’attendais pas à ça !”

Mesure de la divergence entre la
prédiction et la réalité

critères informatiques

surprise maximization

Mesure de la distance entre le prévu et le perçu

learning progress

Mesure de l’amélioration de la prédiction

information gain

Mesure de l’amélioration de sa confiance de la prédiction

Mais comment modéliser ceci ?

Nouveauté

mesure le degré de rareté de l'état rencontré

critères informatiques

Information gain over density model

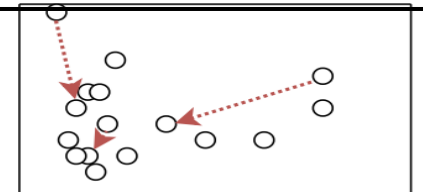
K plus proche voisins

Gain d'information à visiter un état mesuré

- par le décompte de passage dans l'état
- ou via l'estimation de la proba d'être dans l'état

Visited states

Distance between a state and its 4-th closest neighbor



“ c’est un lieu inconnu ! “

- Mais comment modéliser ceci ?

skill learning

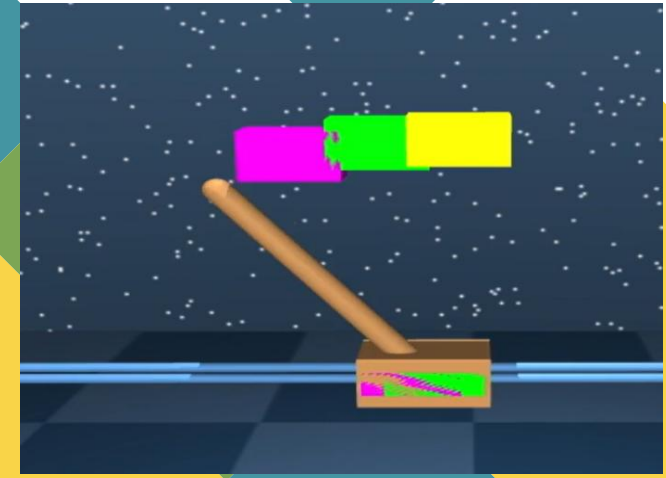
Mesure de l'apprentissage de
compétences réutilisables

critères informatiques

Réussite de sous objectifs préalablement définis

réussite d'objectifs au
hasard

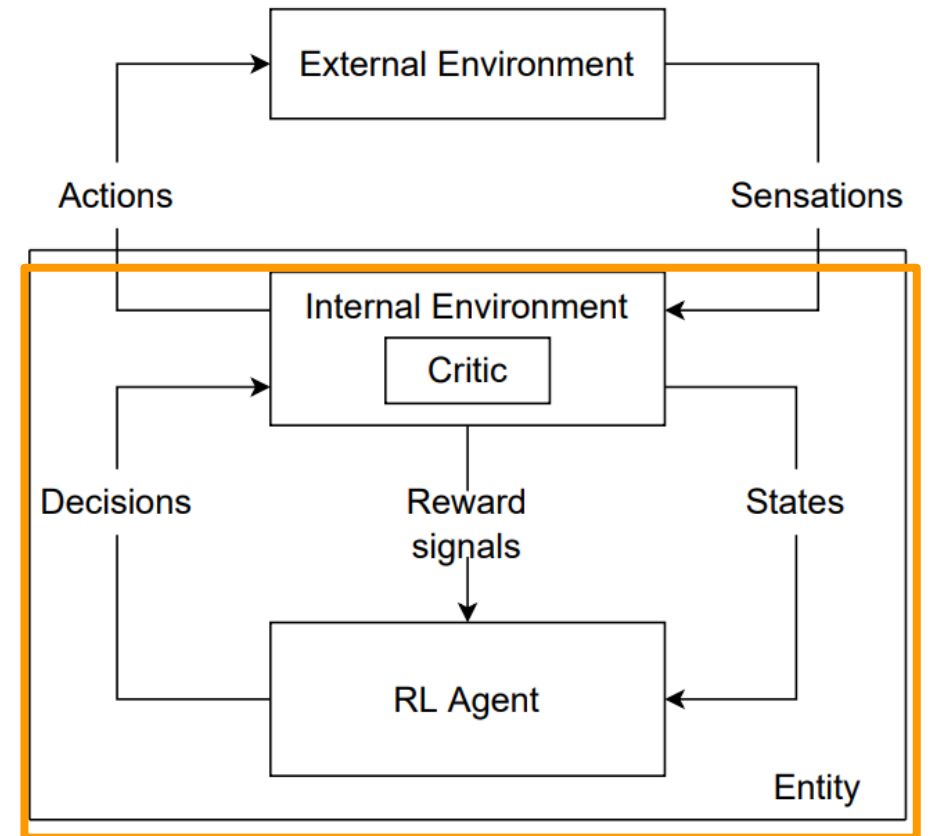
Dans les états déjà
visités / en prenant les
plus rares



- Mais comment modéliser ceci ?

L'ensemble des récompenses (surprise , nouveauté, skill learning, objectif) est **pondéré** dans le circuit interne.

recompense=
F(recompense_interne,
recompense_externe)



Exemples d'utilisation Montezuma's Revenge



sparse reward

nécessité
d'exploration et
d'apprentissage

Les méthodes de motivations
intrinsèques sont utilisées pour
résoudre le jeu.

Exploration et apprentissage

mais reste encore loin derrière
d'autres méthodes
d'apprentissage

zoom sur quelques défis actuels !

Intelligent Adaptive Curiosity

La méthode CIM

Un enfant apprend progressivement par rapport à ce qu'il connaît déjà et de ses capacités motrices



Se retourner
S'asseoir



Ramper



Marcher



La motivation d'un enfant n'est pas guidée par un objectif mais par



Sa

Comment choisir la prochaine expérience pour que le robot apprend le plus ?



Verena Hafner

Professor of Adaptive Systems
(Humboldt-Universität)
Head of the Adaptive Systems
Group
Phd Artificial Intelligence Lab



Frederic Kaplan

Ex-Sony CSL Paris, puis
professeur à l'EPFL, Digital
Humanities Lab
Cofondateur d'OZWE, interfaces
gestuelles



Pierre-Yves Oudeyer

Directeur de recherche Inria
Ex-Sony CSL

Approche 1 ($\max(\delta)$)

Les endroits où je me trompe le plus sont ceux où il me reste le plus à apprendre.

- ✓ Efficace dans une simulation sans bruit
- ✗ Reste bloqué sur des bruits irréductible
- ✗ Erreur élevée $\neq$ quelque chose à apprendre



Approche 2

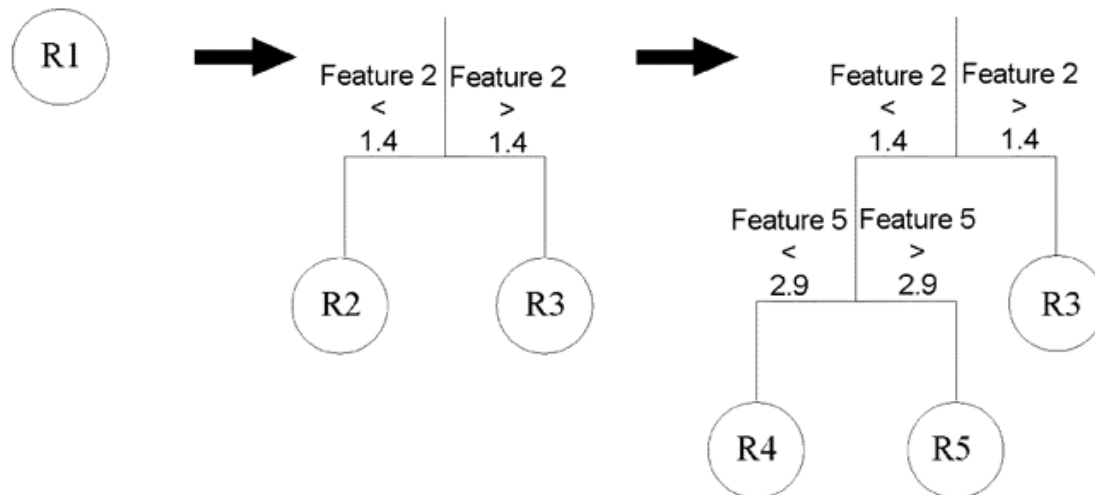
Compare l'erreur moyenne récente et l'erreur moyenne attendue.

- ✓ Plus performant face au bruit (ex: mur)
- ✗ comparaison sur une seule courbe
- ✗ Tâche dure => Tâche facile = Chute de l'erreur



Approche 3 (IAC)

- Stocke toutes les expériences.
- Segmente son espace sensorimoteur en branches
- Chaque branche est dotée de son propre expert.
- Le robot choisit l'action avec l'erreur prévue la plus basse



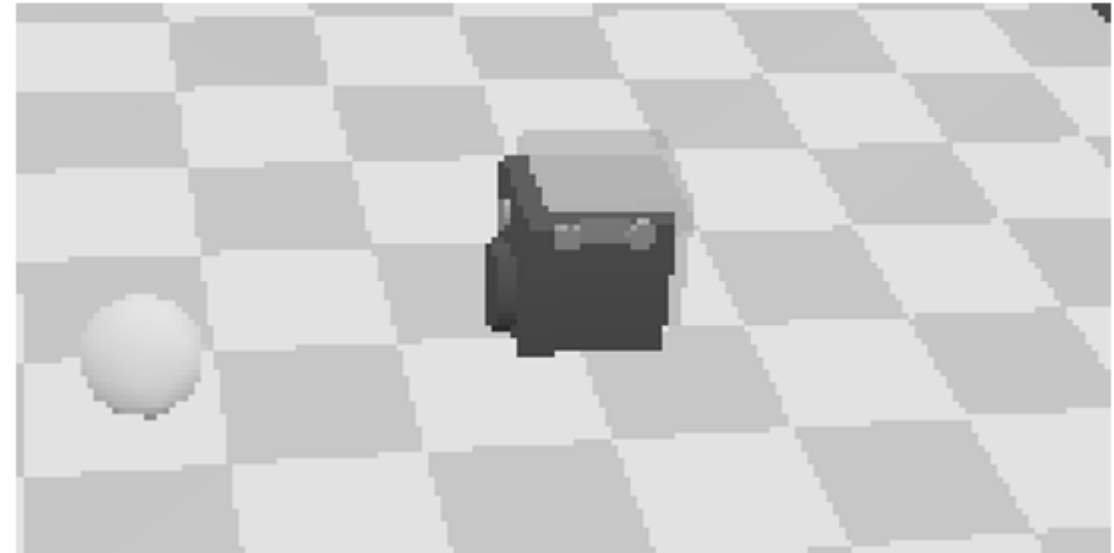
Objectif : Maintenir le progrès d'apprentissage au maximum.

Pour chaque branche

Fonctionnement de l'IAC

Un robot émet un son. Selon sa fréquence, le jouet :

- f1 : bouge au hasard => **imprévisible**
- f2 : s'arrête => **difficile mais apprenable**
- f3 : saute sur le robot => **facile**

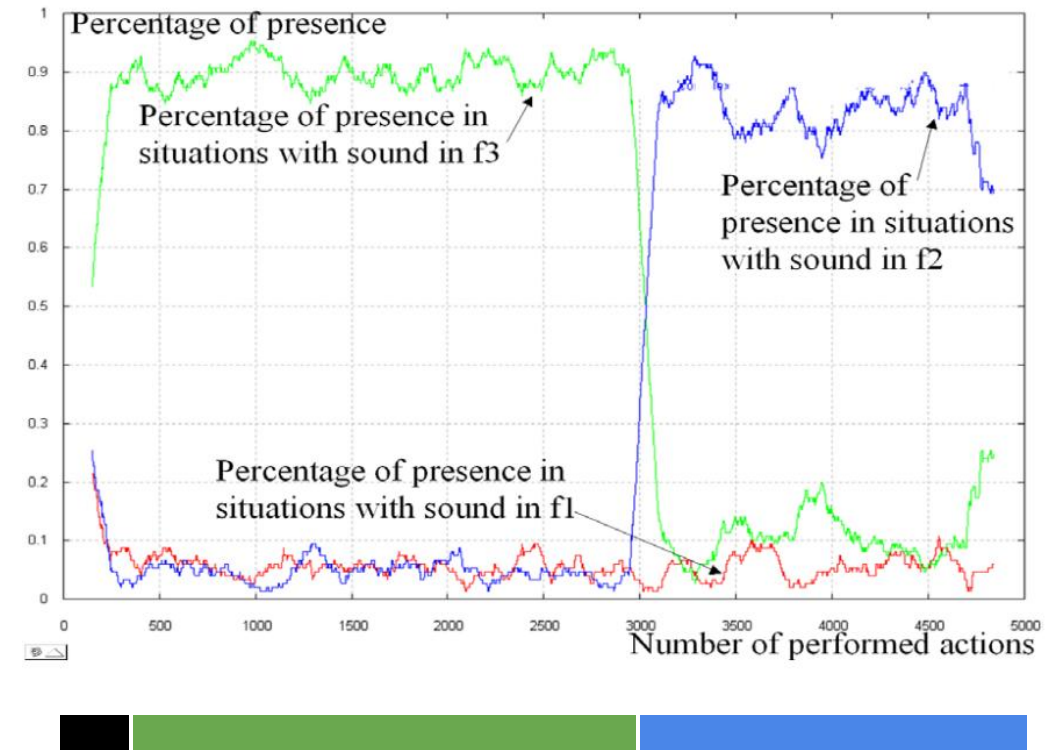


But : **ignorer l'imprévisible**, **épuiser le facile**, **se concentrer sur le difficile mais apprenable**

Le robot organise seul son apprentissage

Part du temps passé dans la zone :

- Vers **0 à 250** : Exploration au hasard
- De **250 à 3 000** : Apprend la zone facile
- Après **3 000** : Facile maîtrisé, bascule vers le difficile



Sans connaître les zones, le robot passe du facile au difficile et ignore l'impossible.

La curiosité est un vilain défaut



Les limites de la motivation intrinsèque classique :

- **Déraillement** : l'agent poursuit la nouveauté au détriment de l'objectif
- **Myopie** : les zones a priori peu stimulantes mais nécessaires sont délaissées
- **Imprévisibilité** : pas de garantie de progression vers le but

Enjeu : comment réguler cette curiosité ?

La méthode CIM



Pendant l'entraînement :

Diversité forcée : récompenser les mouvements les plus différents possibles pour une tâche

Résultat : un répertoire de comportements riche

Pendant l'exploration :

Curiosité pondérée : plus l'agent se rapproche du but, moins l'exploration est intéressante

Résultats



- **Meilleurs couverture d'états**
 - 1024 cases explorées contre 697 avec CIC sur robot ants
- **Efficacité d'échantillonnage**
 - 20x moins d'échantillon que LSD
- **Elimination des distraction**

Limites

- **Sensibilité à la dimensionnalité**
 - 1042 pour $n = 2$ contre 615 pour $n = 64$ sur robot ants
- **Moins efficace pour les tâches de manipulations**
- **Sensibilité à la fonction de perte**
 - Utiliser 1042 avec NCE-style contre 64 avec MSE classique sur ants
- **Dépendance à un méta-controller**

The Playground Experiment

Ce que le robot peut faire:

- Tourner la tête
- Mordre
- Frapper avec la patte

Son environnement:

- Deux jouets : un objet qu'on peut mordre et un objet qu'on peut frapper

Au départ, AIBO ne sait encore rien faire..



Sur 200 essais

Aucune tâche imposée, et pourtant des stades émergent, comme chez un bébé

Scénario complet	34%
Apprend à mordre le bon objet	100%
Apprend à frapper le bon objet	78%
Mord avant de frapper	92%
Mord dans le vide	93%



Conclusion

Le robot ne va pas là où il se trompe le plus, mais là où il progresse le plus.

Points forts	Points faibles
Aller seul du facile vers le difficile	Ne vise que la récompense immédiate
Ignorer les actions impossible à apprendre (aléatoire par ex.)	Ne sait pas organiser ses compétences ni se fixer des buts

D'autres voies de recherches :

- **Introduire une hiérarchie de compétences**
 - Définir et ajuster des compétences de bases servant à des compétences plus élevés
- **choisir une hiérarchie d'objectifs de façon efficiente**
 - introduire une métrique sur le progrès dans une compétence

A vous !



qui a compris quelque chose ?
vous pouvez essayer de résumer ?

A vous !
2 équipes
1 point par réponse

- 1) a quoi ça sert ?**
- 2) qu'est ce que c'est ?**
- 3) quels sont les différents aspects ?**
- 4) est ce que ça marche ?**
- 5) est ce que c'est résolu ?**

Que retenir ?

La motivation intrinsèque se définit par **l'origine du signal de récompense** : la récompense est dans ce cas dérivée de **l'expérience interne de l'agent (surprise, nouveauté, apprentissage)** plutôt que fournie par l'environnement.

Cela permet de **résoudre** des **problèmes avec des récompenses éparses**.

Il y a différentes façon de la **modéliser informatiquement**.

Les résultats semblent prometteurs : L'apprentissage multi-factoriel ne dépend pas d'un seul environnement. Néanmoins, l'ajustement pratique de l'ensemble des paramètres est encore un sujet de recherche.

**A vous !
débat (2 points par arguments)**



Faut il concevoir des IA avec leurs propres motivations ?

A vous ! débat (2 points par arguments)

Pour vous, la motivation intrinsèque, ce serait pas juste du machine learning sophistiqué ?

Débat:



Au vu des résultats, pourquoi on a pas la solution miracle ?