

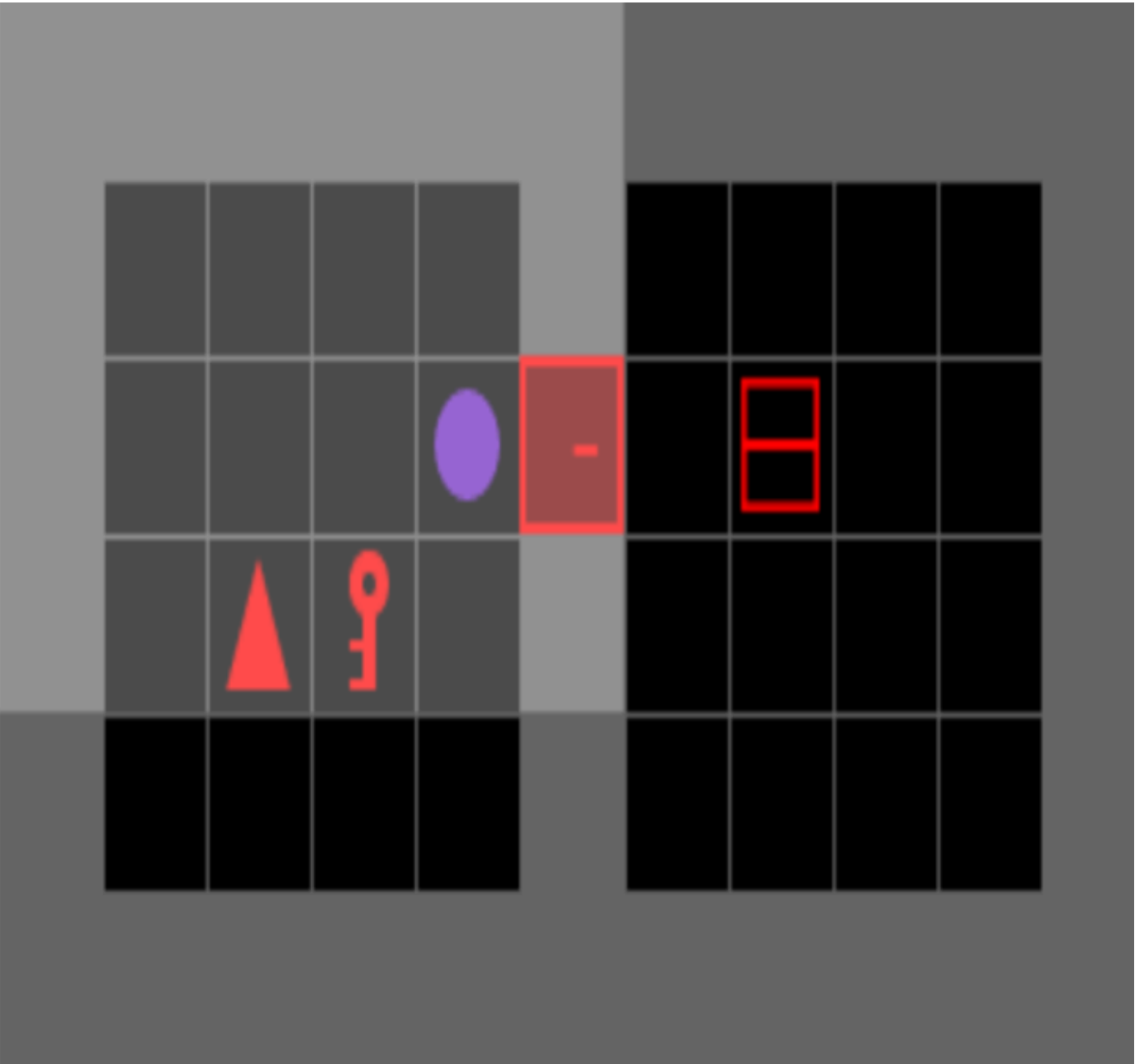
# Reinforcement Learning: BlockedUnlockPickup

github: [https://github.com/suprawall/MiniGrid\\_BlockedUnlockPickup](https://github.com/suprawall/MiniGrid_BlockedUnlockPickup)

- git clone
- cd Minigrid: pip install -e .
- cd .. (dans le dossier clone)
- executer projet.py

Valentin Hesters  
Sara Firoud

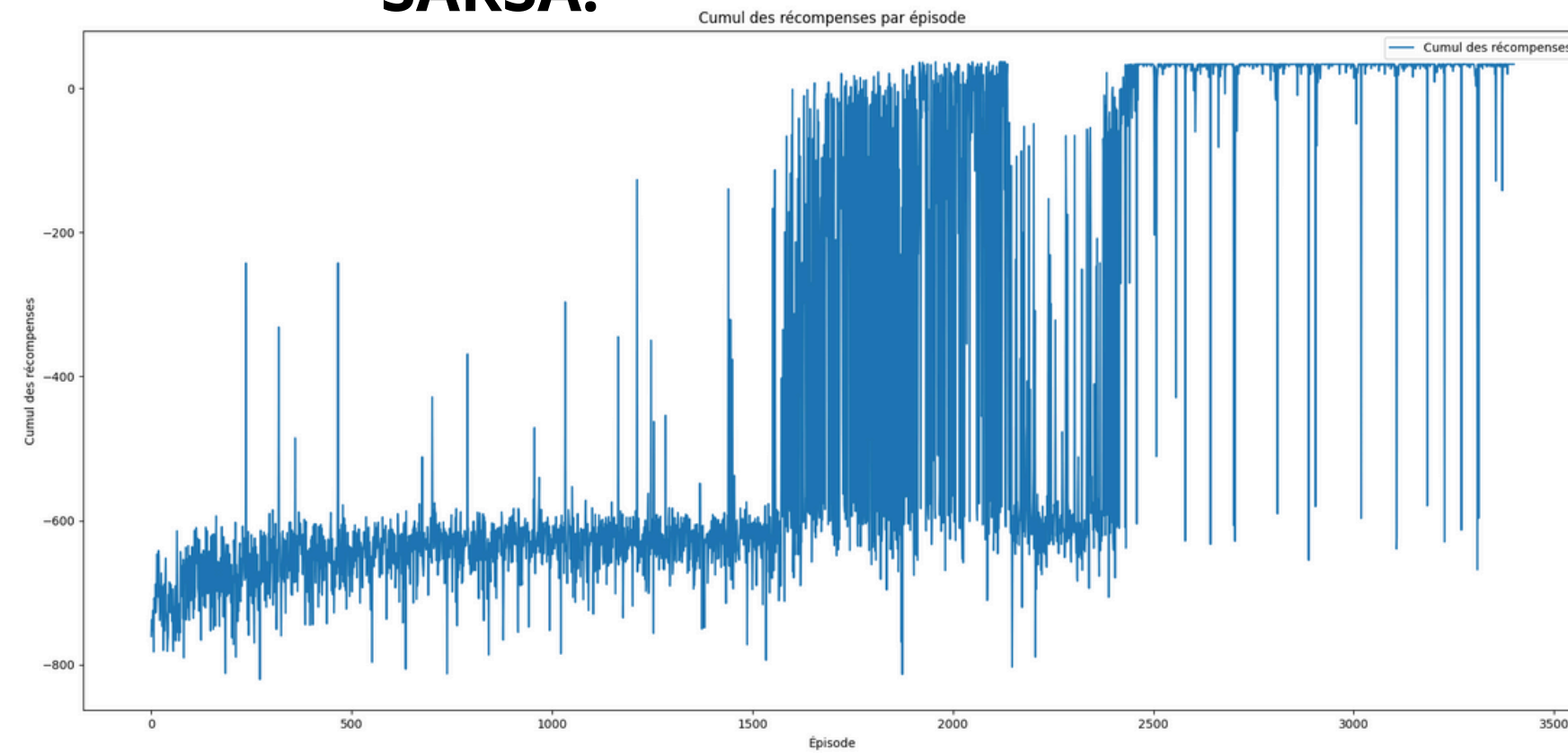
---



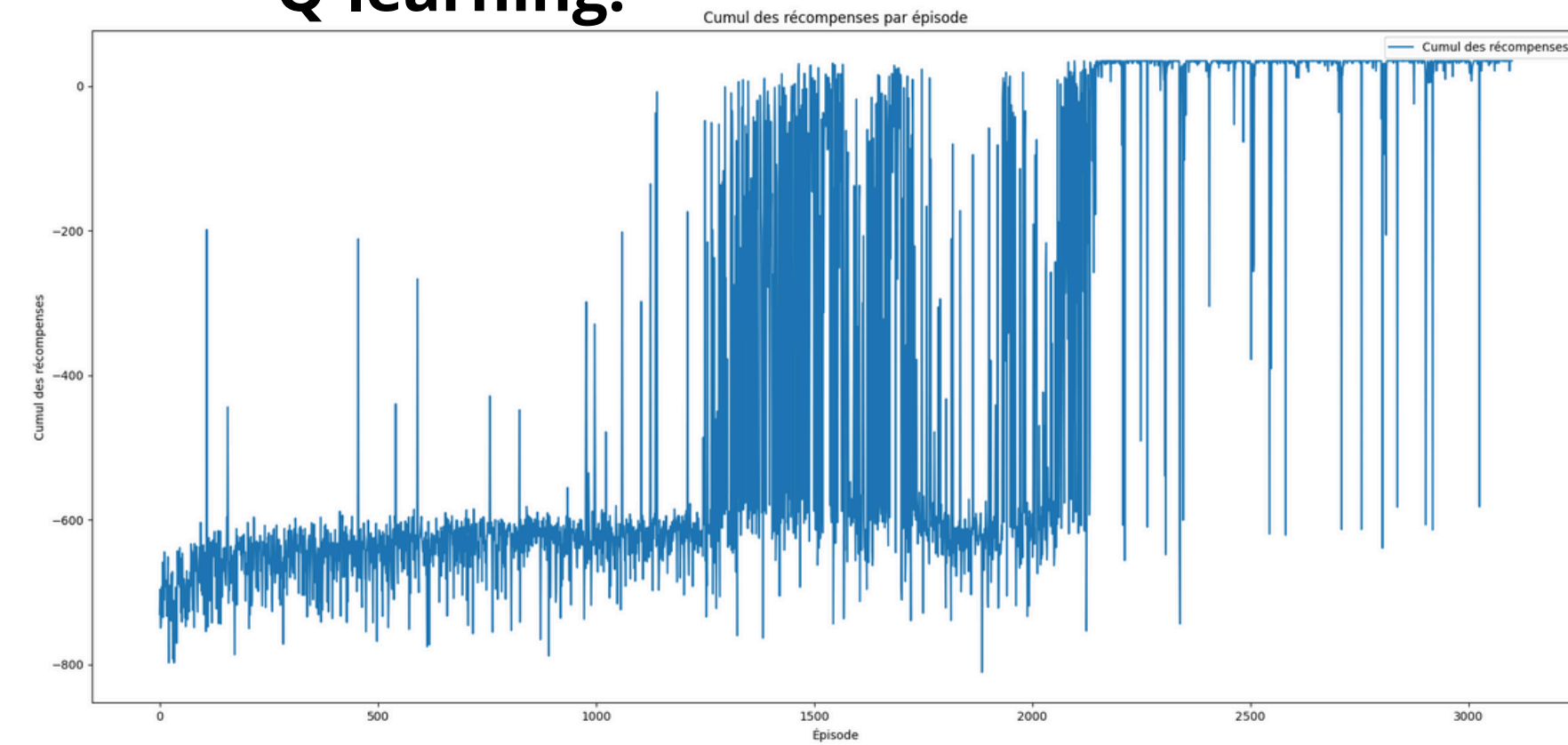
$$\text{Reward}(P) = P + \frac{10 * (\text{max\_steps} - \text{steps})}{\text{max\_steps}}$$

	récompense	Comportement indésirable	Solution
mouvement	-1		
pick-up box	reward(15)		
toggle locked door	reward(4)	locked/toggle en boucle	-2 si elle a déjà été ouverte
pick-up ball/key	reward(2)	pickup en boucle	-2 si : .déjà pris par le passé .on a déjà un item .pickup le vide
drop	-2	éviter de drop sans avoir d'item	

# SARSA:

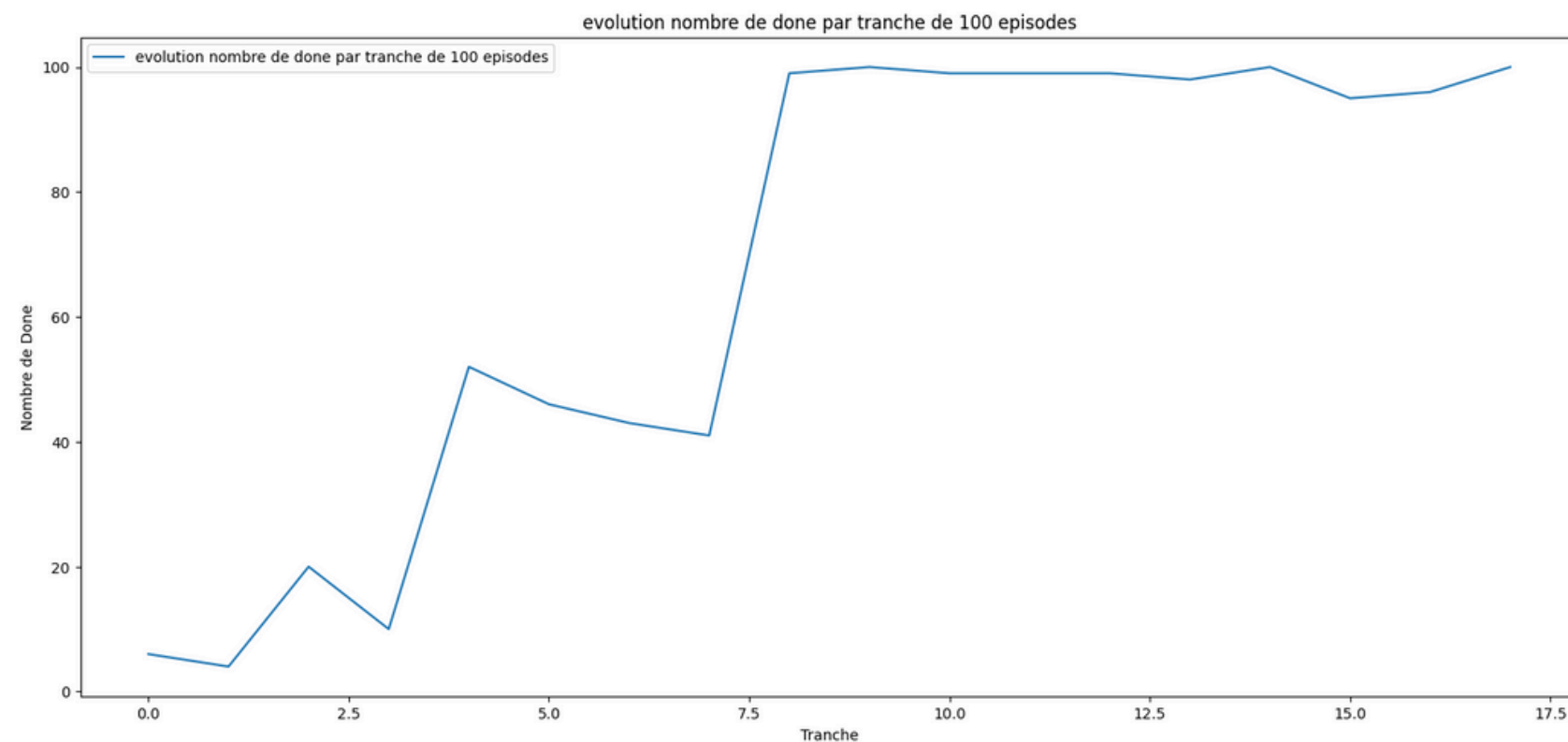


# Q-learning:



```
===== Debut de l'entrainement avec Q-learning sur la seed 5818 =====  
Episode 100, Q-table size: 4608, nombre de done: 3 temps: 6.708513021469116  
Max Reward: -296.5400000000002 sur l'épisode 80
```

```
-----  
Episode 1300, Q-table size: 8485, nombre de done: 98 temps: 0.3869051933288574  
Max Reward: 58.42 sur l'épisode 1200
```



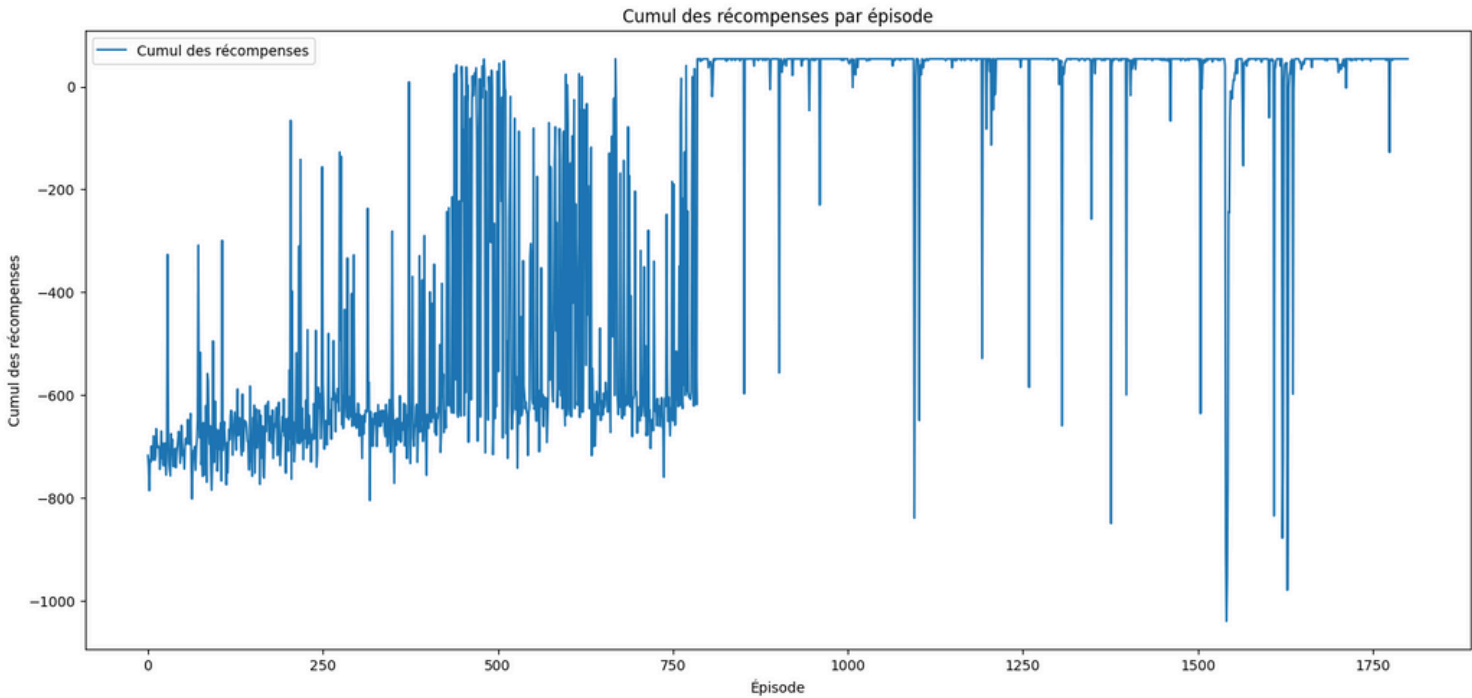
- Entrainement sur plusieurs seeds lourd (20 seeds  $\sim\sim$  500Mo)
- Mauvais en généralisation: trop de combinaisons possibles

**Nouveau cerveau:** [ ‘ball’: (dist, dir), ‘key’: (dist, dir), ‘door’: (dist, dir), ‘box’: (dist, dir, state),  
                  ‘is\_wall\_in\_front’: 0 || 1]

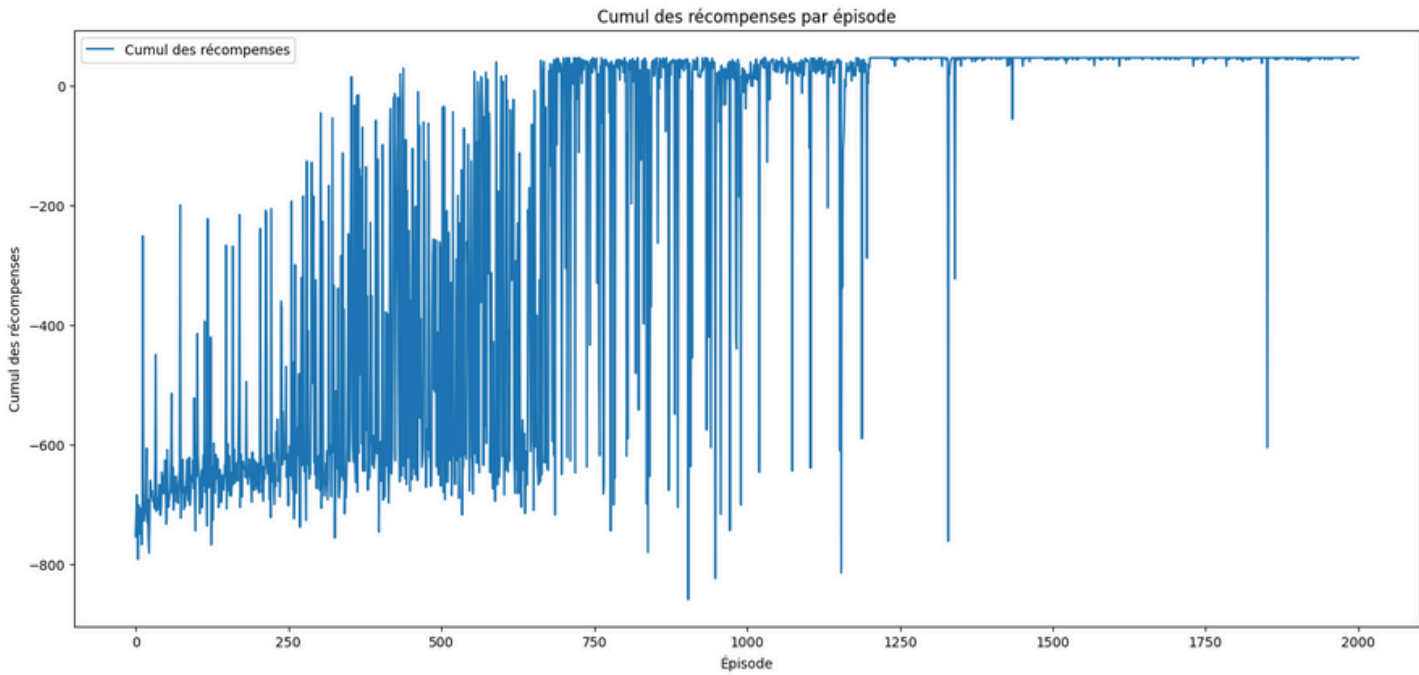
-----

dist: 0 à 9 (inf quand non visible)    dir: 0: tout droit, 1: à gauche, 2: à droite    state: 0: locked, 1: closed, 2: open

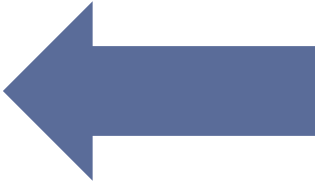
ancien:



nouveau:



Max step Type de cerveau	50	100	500	1000
Nouveau	51%	72%	90%	96%
Ancien	0%	0%	2%	8%



**Environnement jamais  
vu par l'agent**

# Perspectives:

- Cerveau plus complet: trop d'états identiques sur des configurations différentes
- Prise en compte des couleurs: nouveaux environnements avec des clefs pièges de différentes couleurs
- Modifier le système de reward: moins précis
- Aggrandir les pièces et entrainer sur l'exploration

