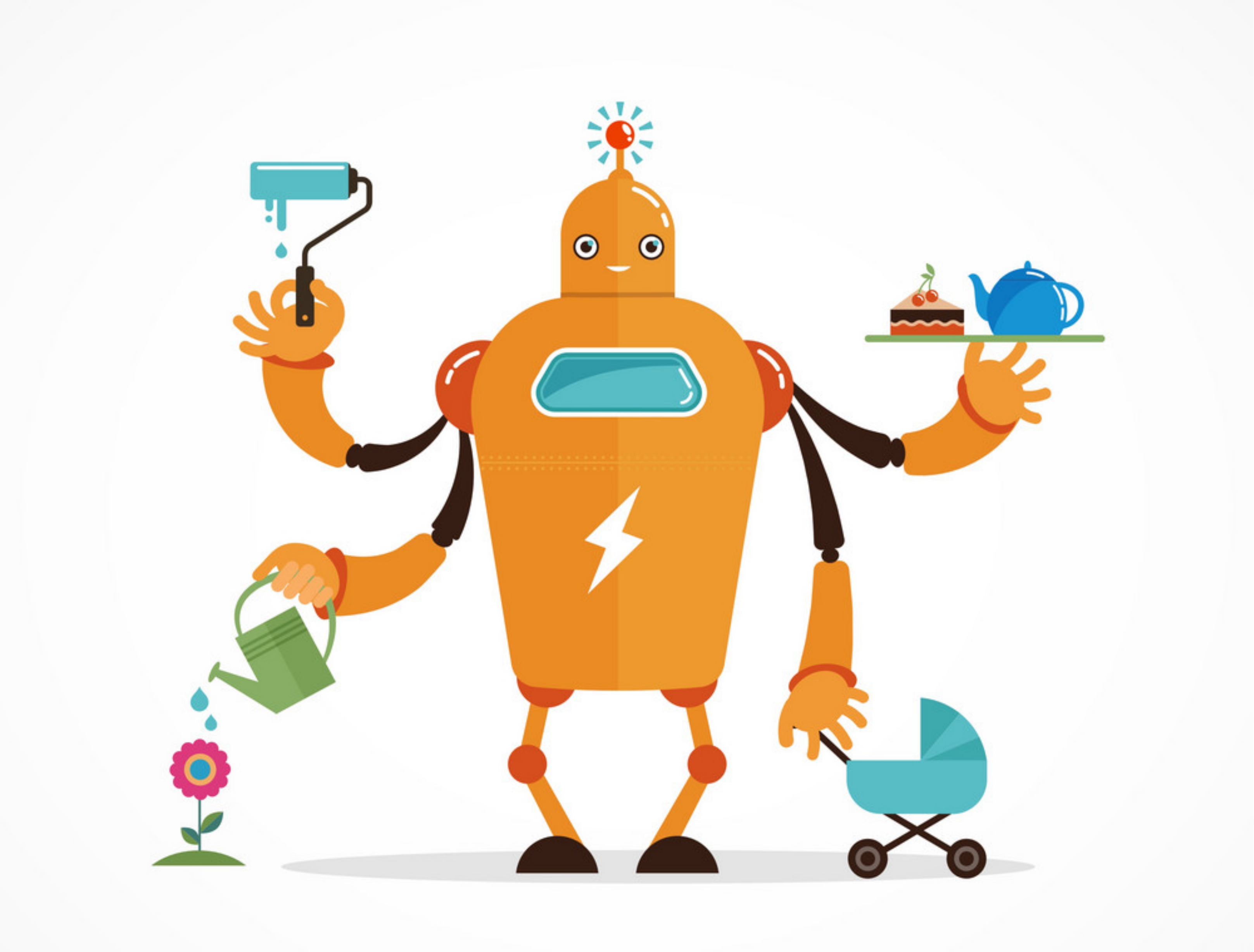
Better Transfer Learning with Inferred Successor Maps

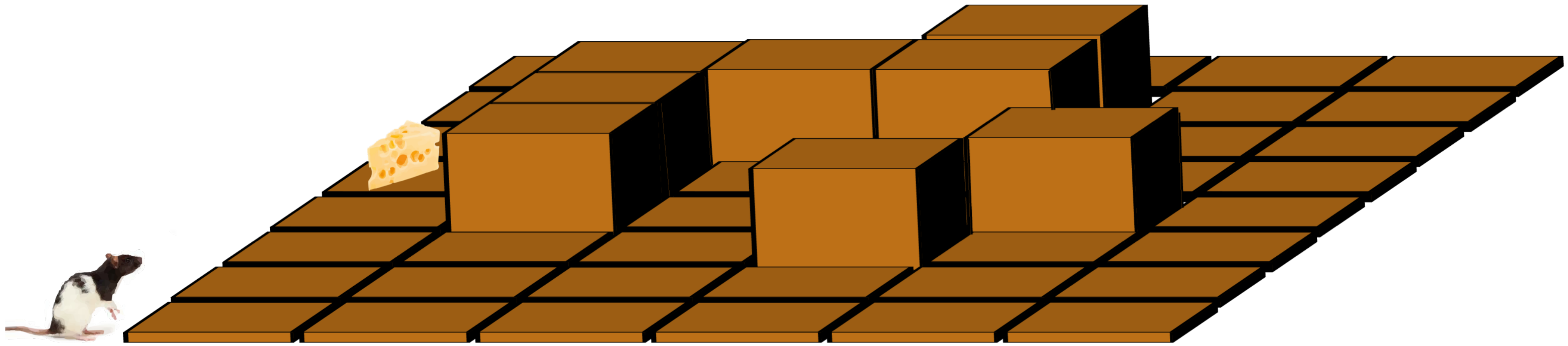
Tamas Madarasz[1,2], Tim Behrens[1,2]

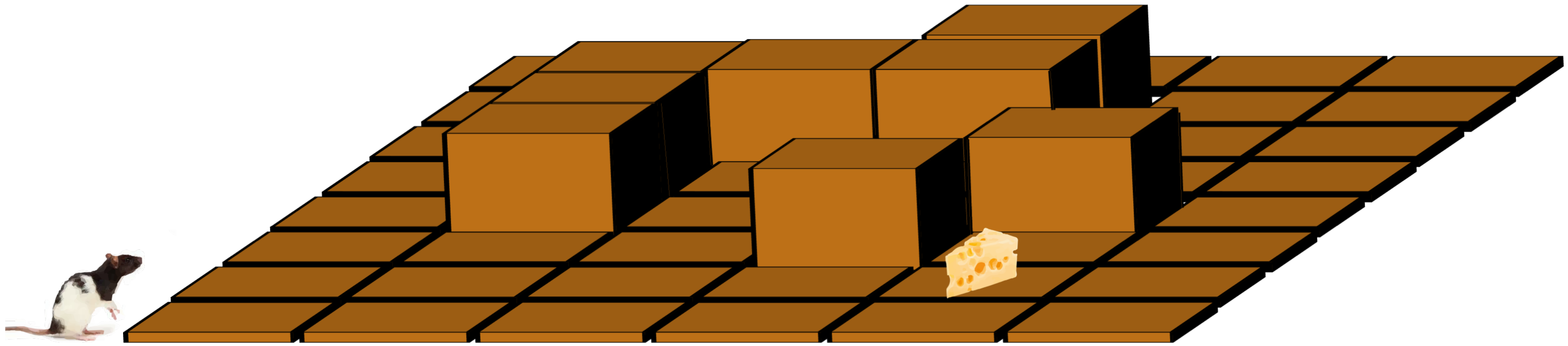
arXiv:1906.07663

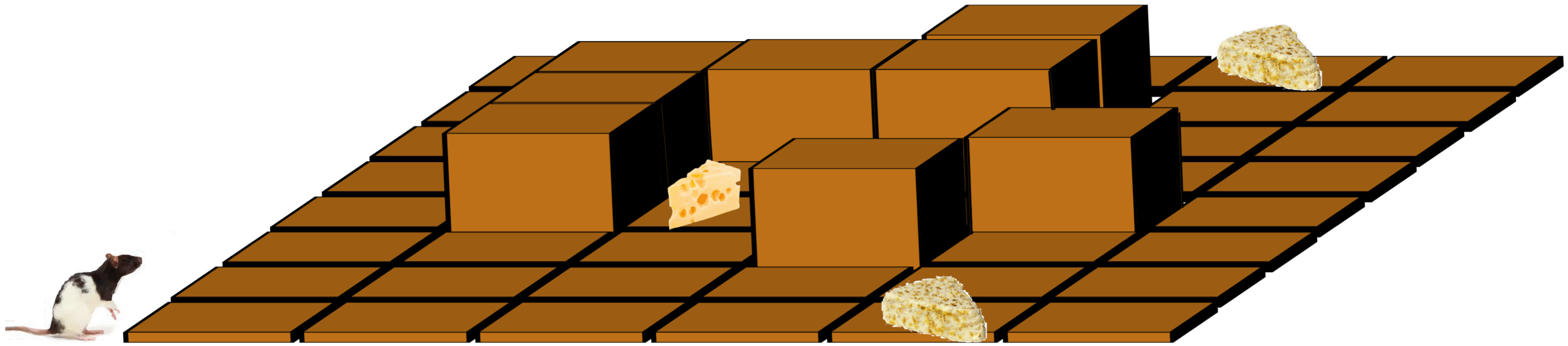
Spotlight NeurIPS 2019

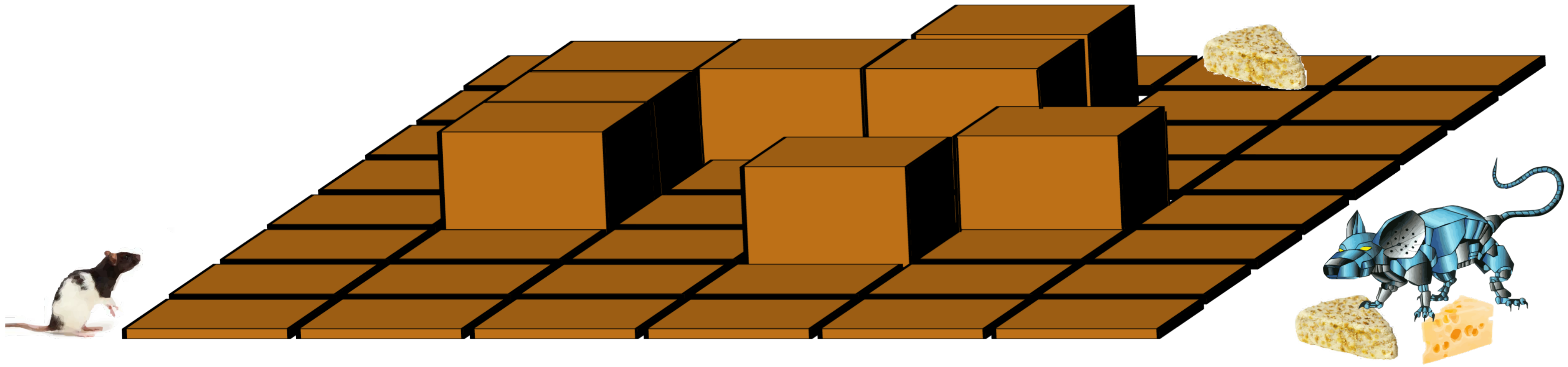
1: University of Oxford
2: UCL

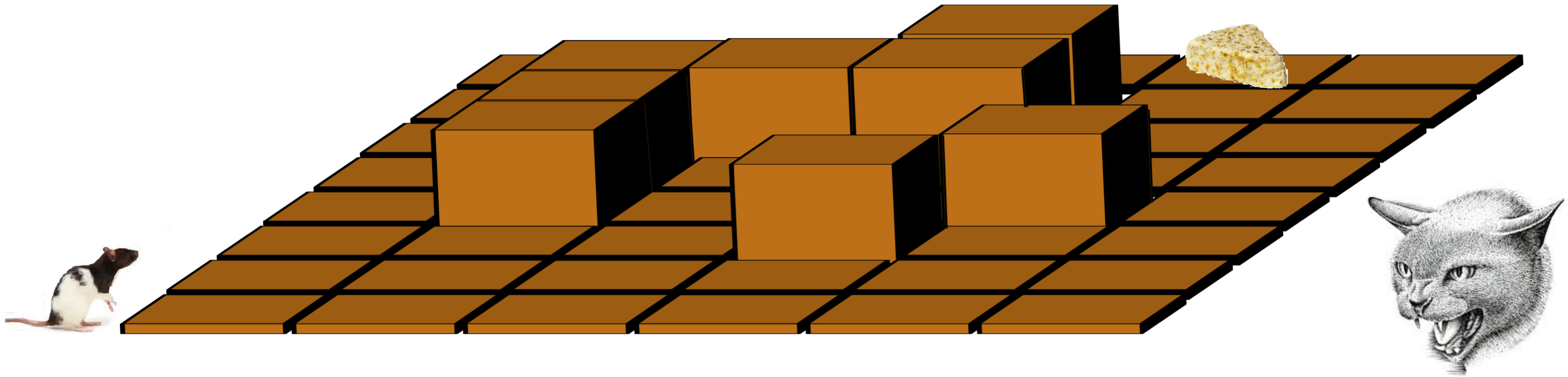


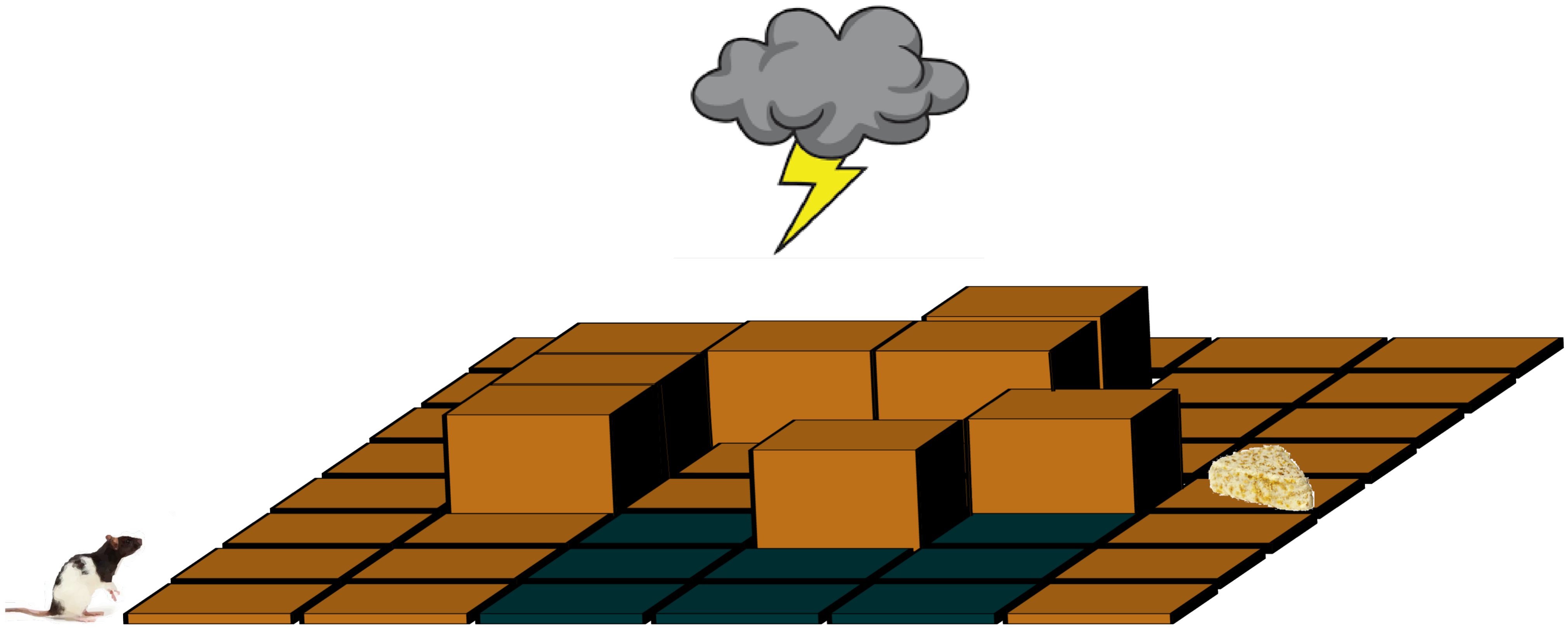




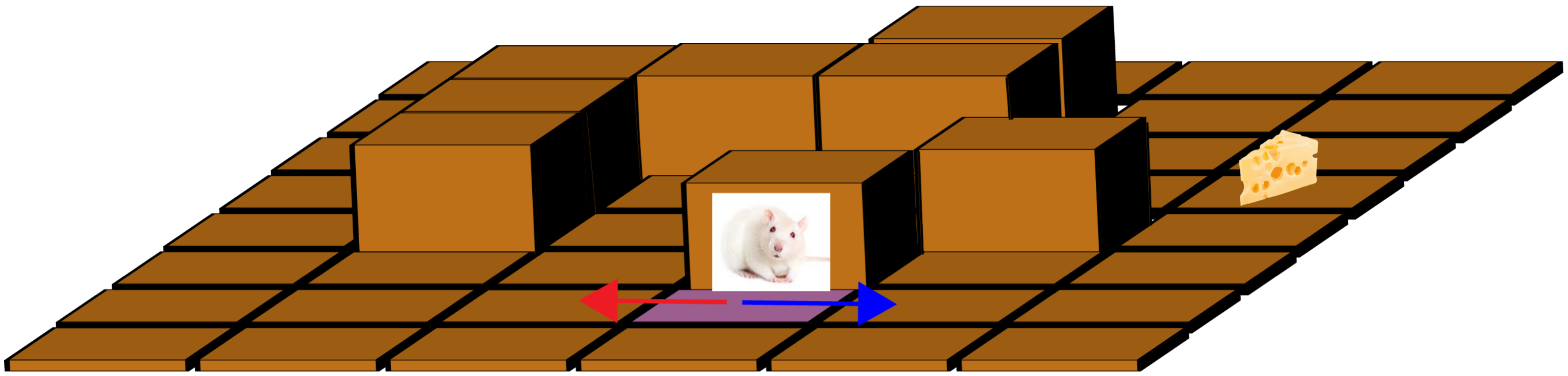




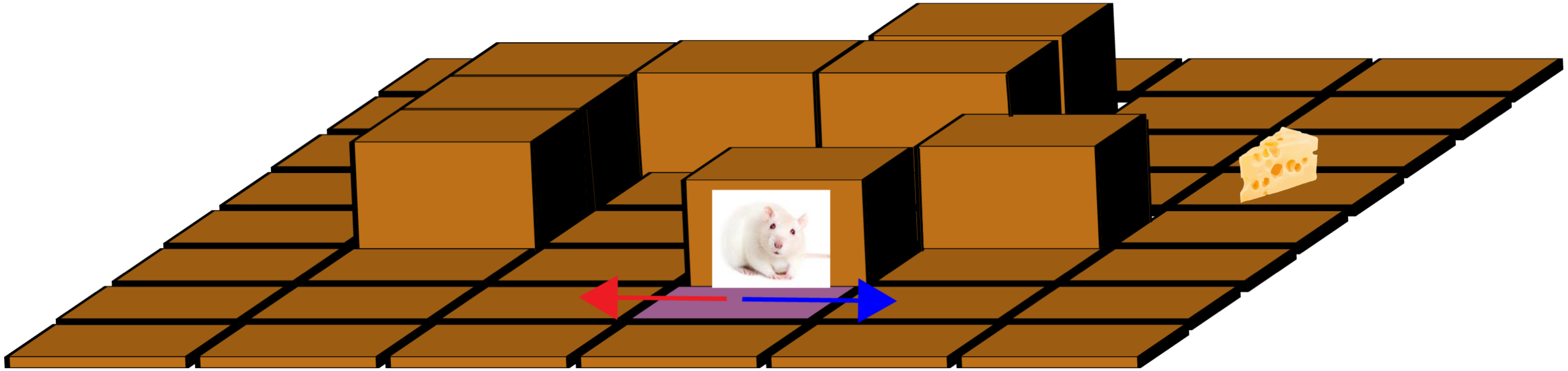




$$Q_t^\pi(s,a) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} \middle| s_t = s, a_t = a\right]$$

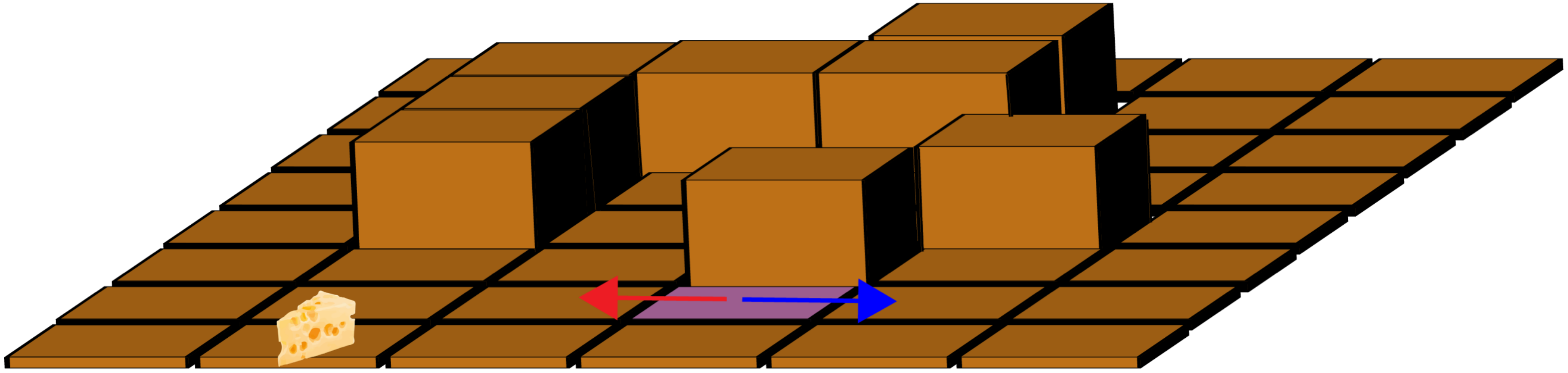


$$Q_t^{\pi}(s,a) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} \middle| s_t = s, a_t = a\right]$$



$Q^{\pi}(s, a_1)$ $\boxed{Q^{\pi}(s, a_2)}$

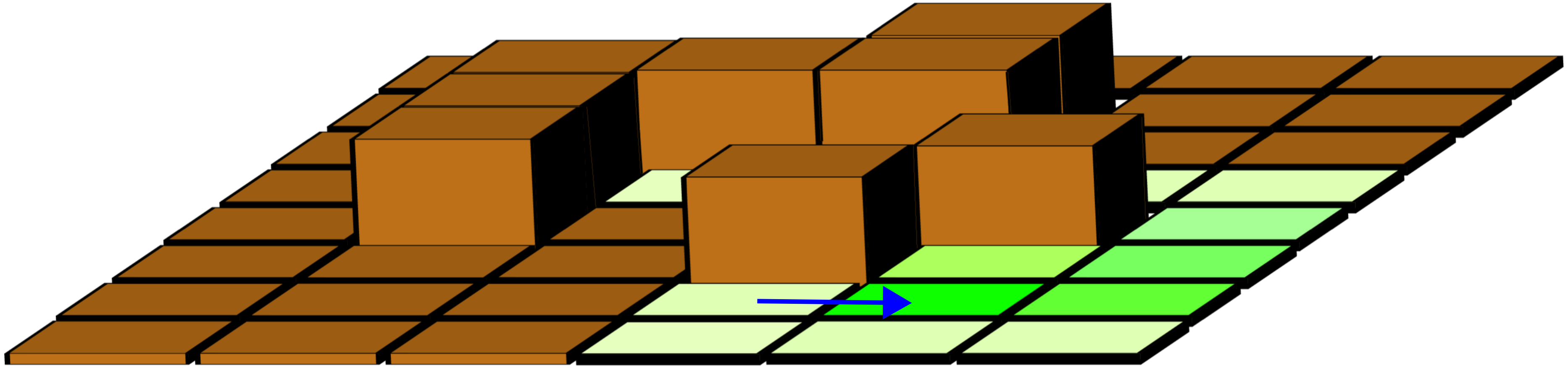
$$Q_t^{\pi}(s,a) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} \middle| s_t = s, a_t = a\right]$$



$Q^{\pi}(s, a_1)$ $Q^{\pi}(s, a_2)$

The successor representation (SR)

$$M_t^\pi(s,a,s') = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \mathbb{I}_{(s_{t+k+1}=s')} \middle| s_t = s, a_t = a\right]$$

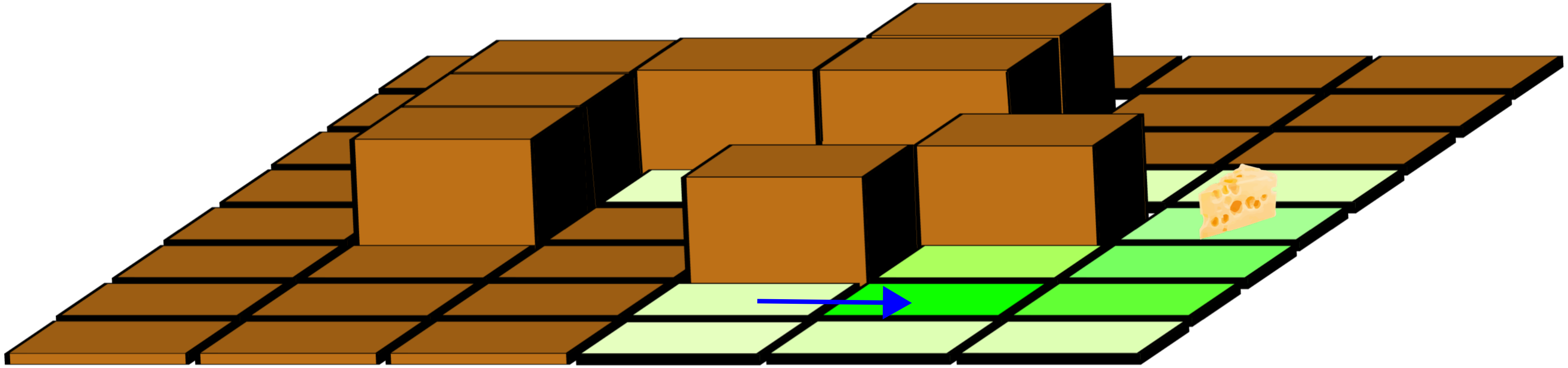


$M^\pi(s,a,:)$

Dayan, 1993
Neural Computation

The successor representation (SR)

$$M_t^{\pi}(s, a, s') = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \mathbb{I}_{(s_{t+k+1}=s')} \middle| s_t = s, a_t = a\right]$$

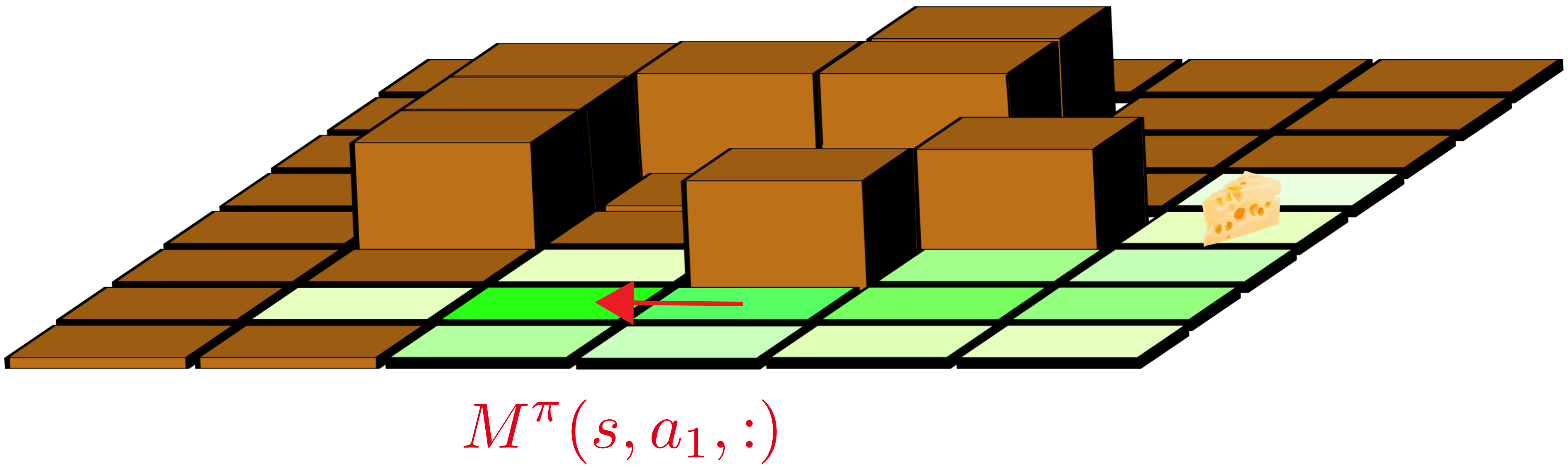


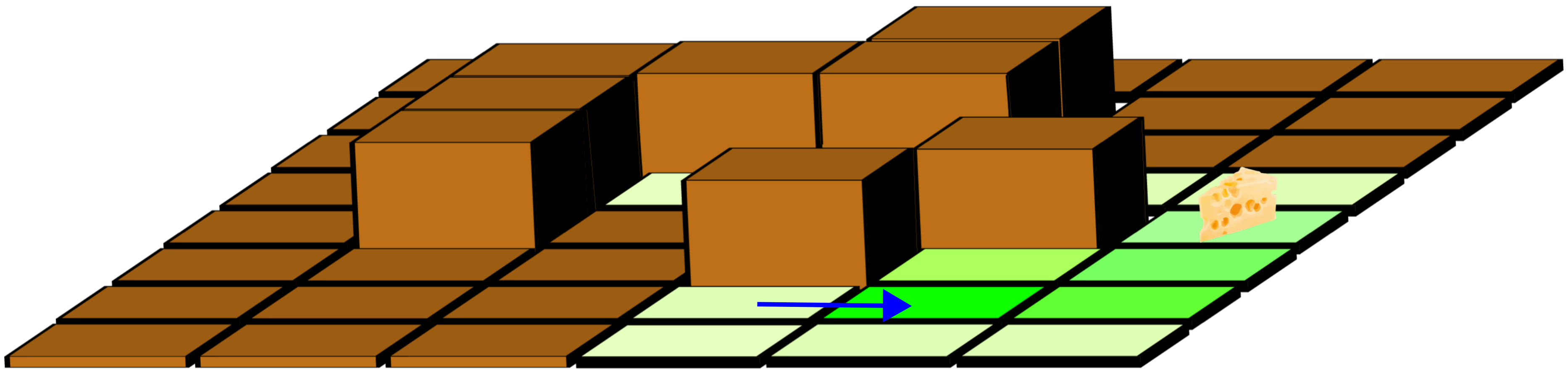
$M^{\pi}(s, a, :)$

$$Q^{\pi}(s, a) = \sum_{s'} M(s, a, s') \cdot \mathbf{w}(s')$$

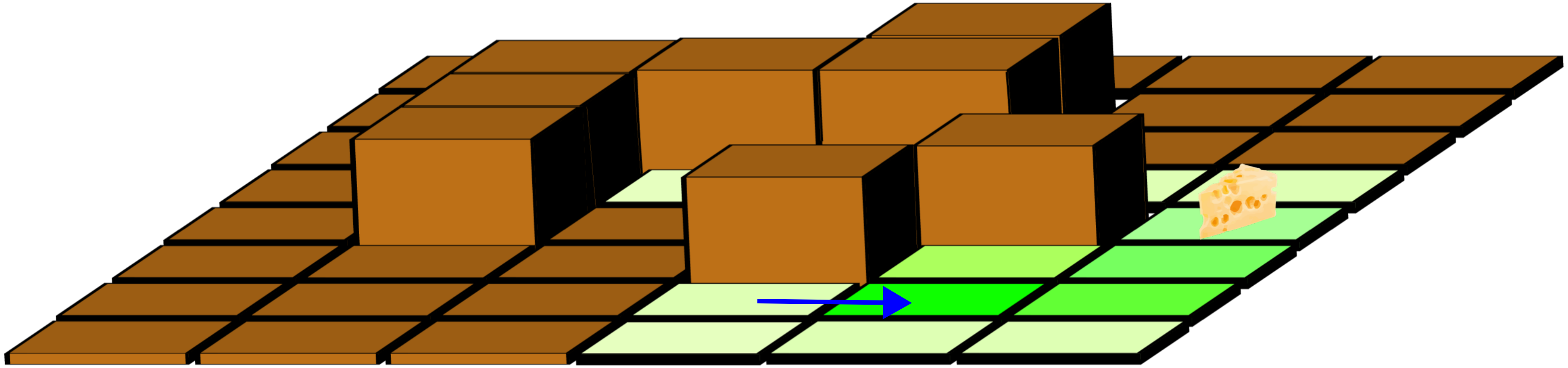
reward function

Dayan, 1993
Neural Computation

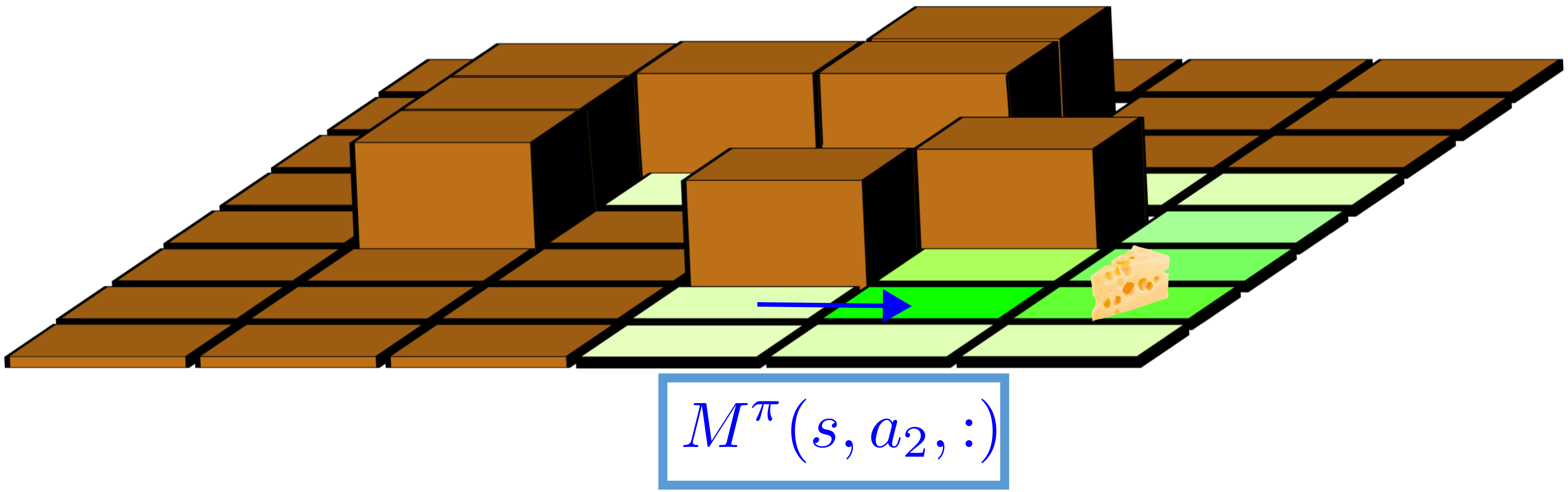

$M^\pi(s, a_1, :)$

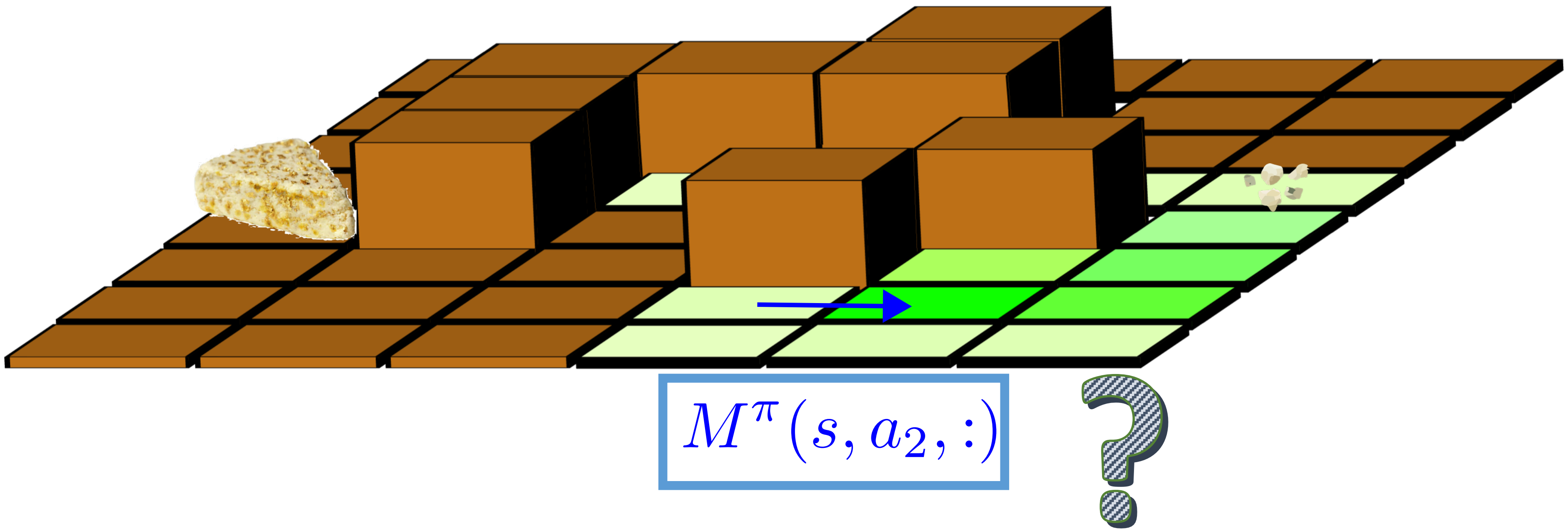


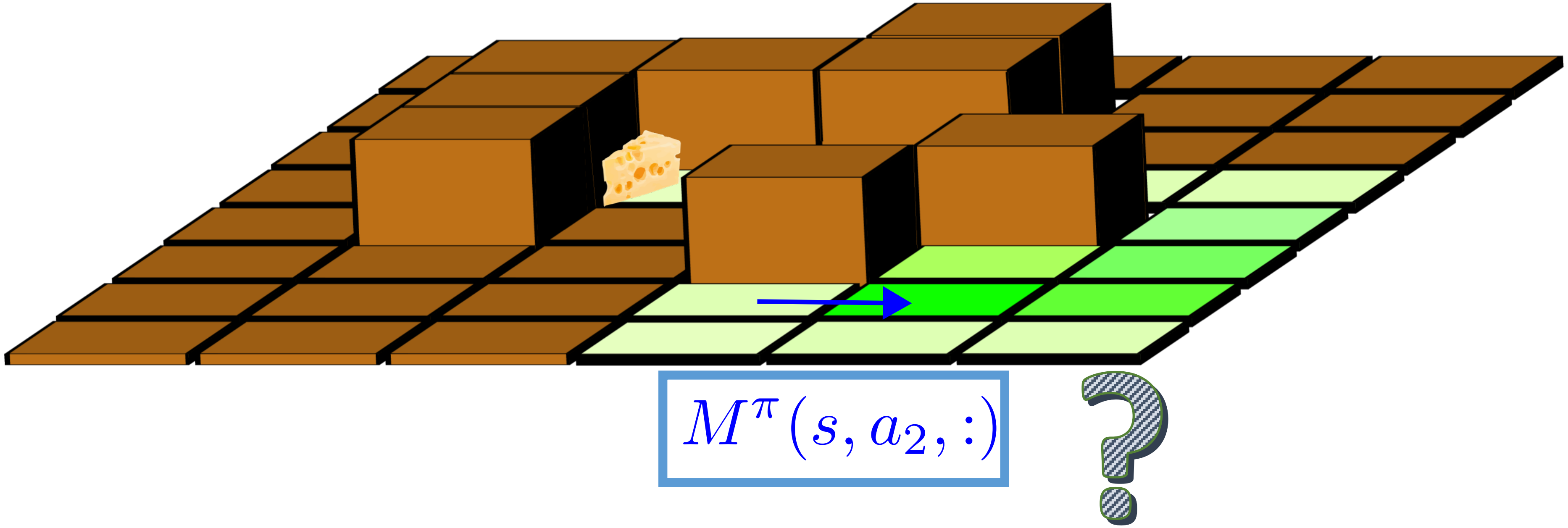
$M^{\pi}(s, a_2, :)$



$M^\pi(s, a_2, :)$


$M^\pi(s, a_2, :)$


$M^{\pi}(s, a_2, :)$


$M^{\pi}(s, a_2, :)$

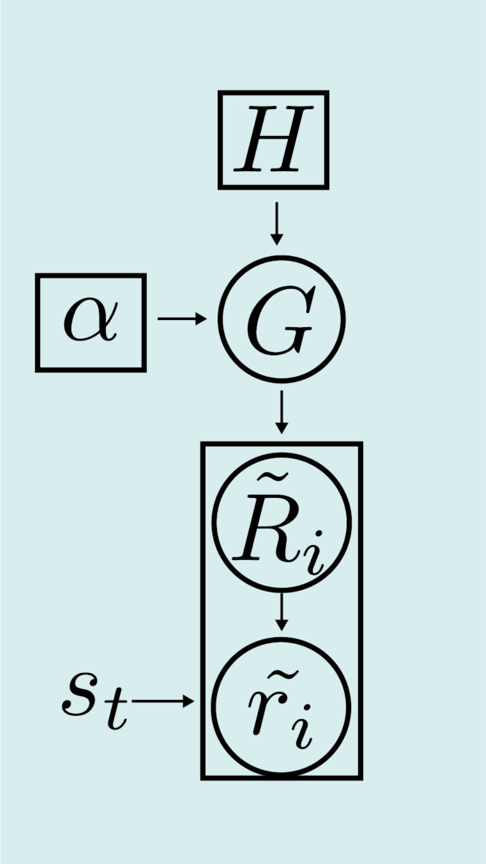
Main approach

- Cluster tasks and try to map current task to the cluster such that SR is easiest to adapt

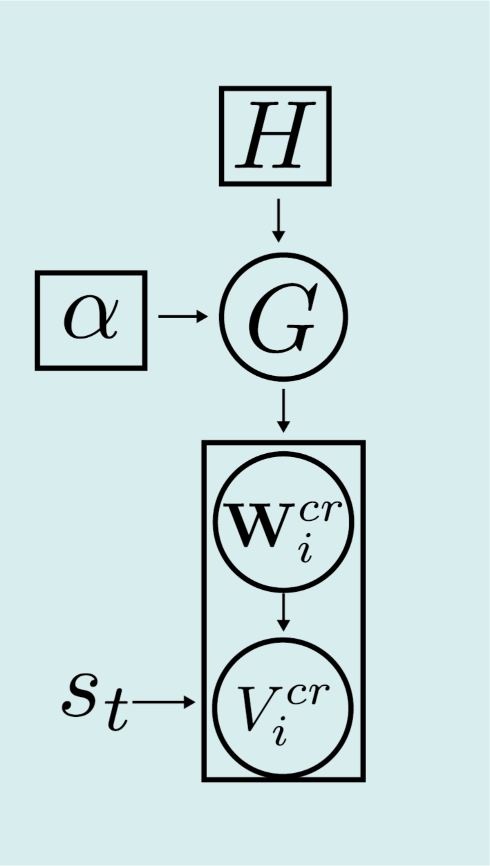
- Use the SR's flexibility to approximate the optimal value function

Wilson et al. 2007, *ICML*
Lazaric and Ghamazadev 2010*, ICML*
Finn et al. 2017, *ICML*

Generative model over reward functions



Generative model over reward functions

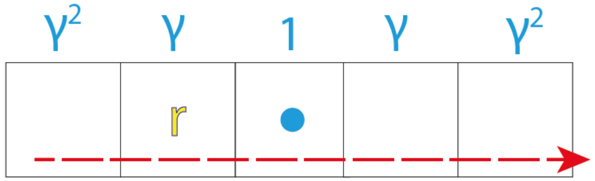


Dirichlet Process mixture model of kernel- smoothed rewards

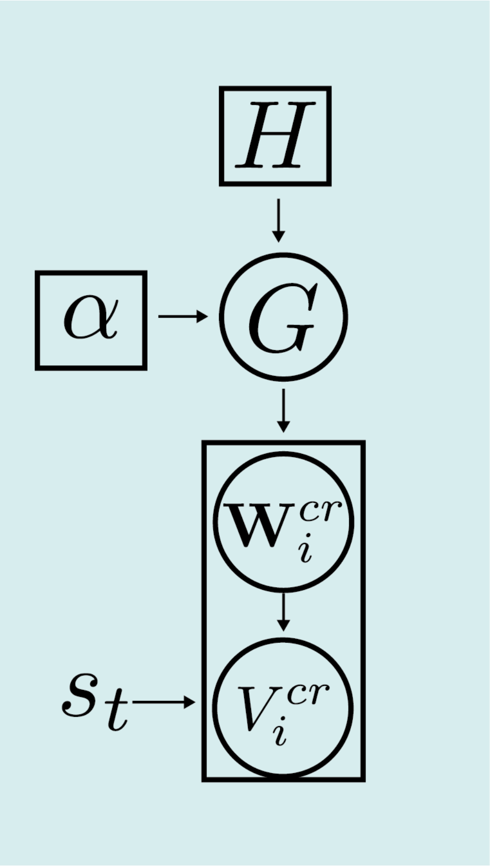
Convolved rewards (CR)

convolution

v_i^{cr} :



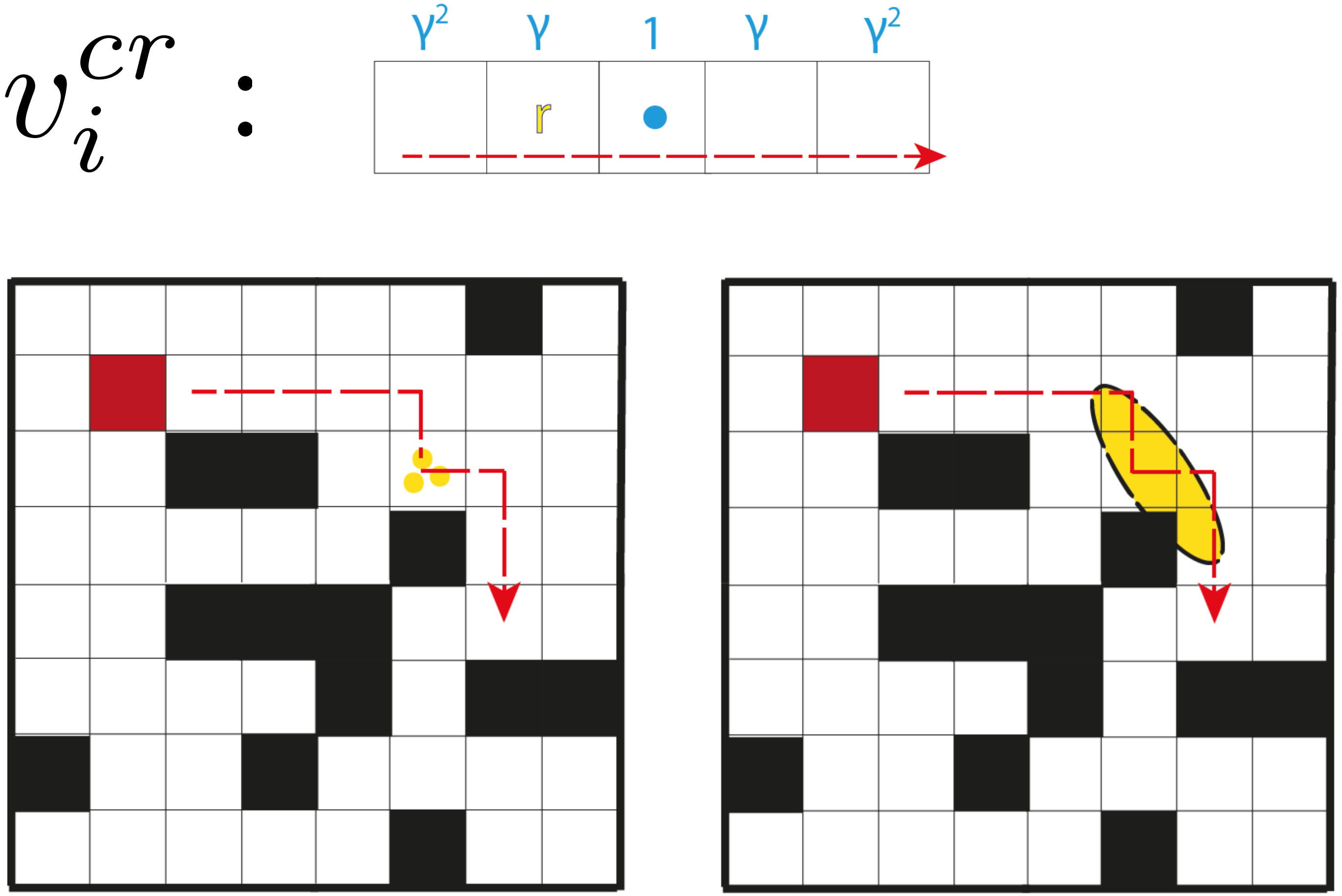
Generative model over reward functions



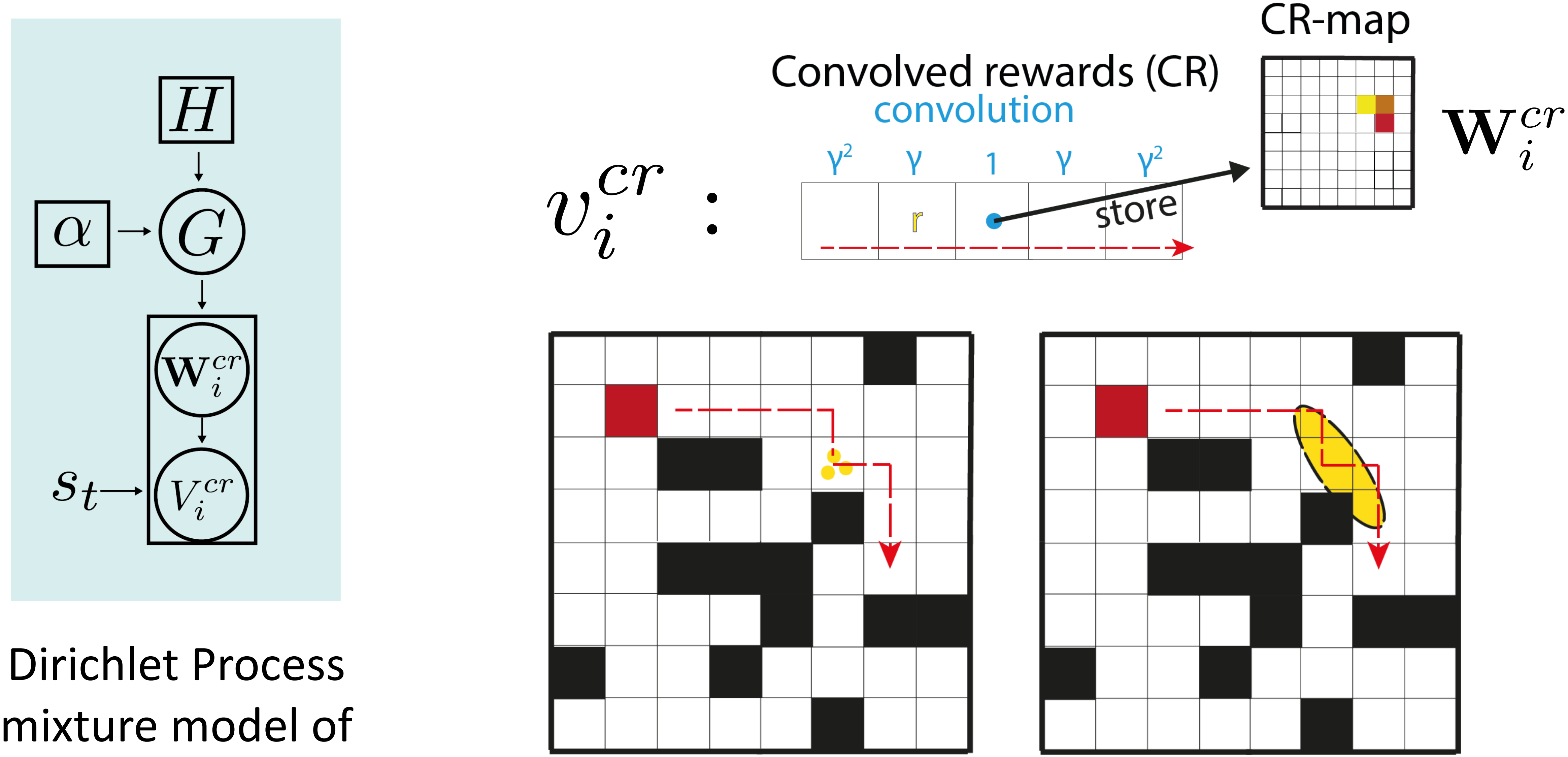
Dirichlet Process mixture model of kernel- smoothed rewards

Convolved rewards (CR)

convolution

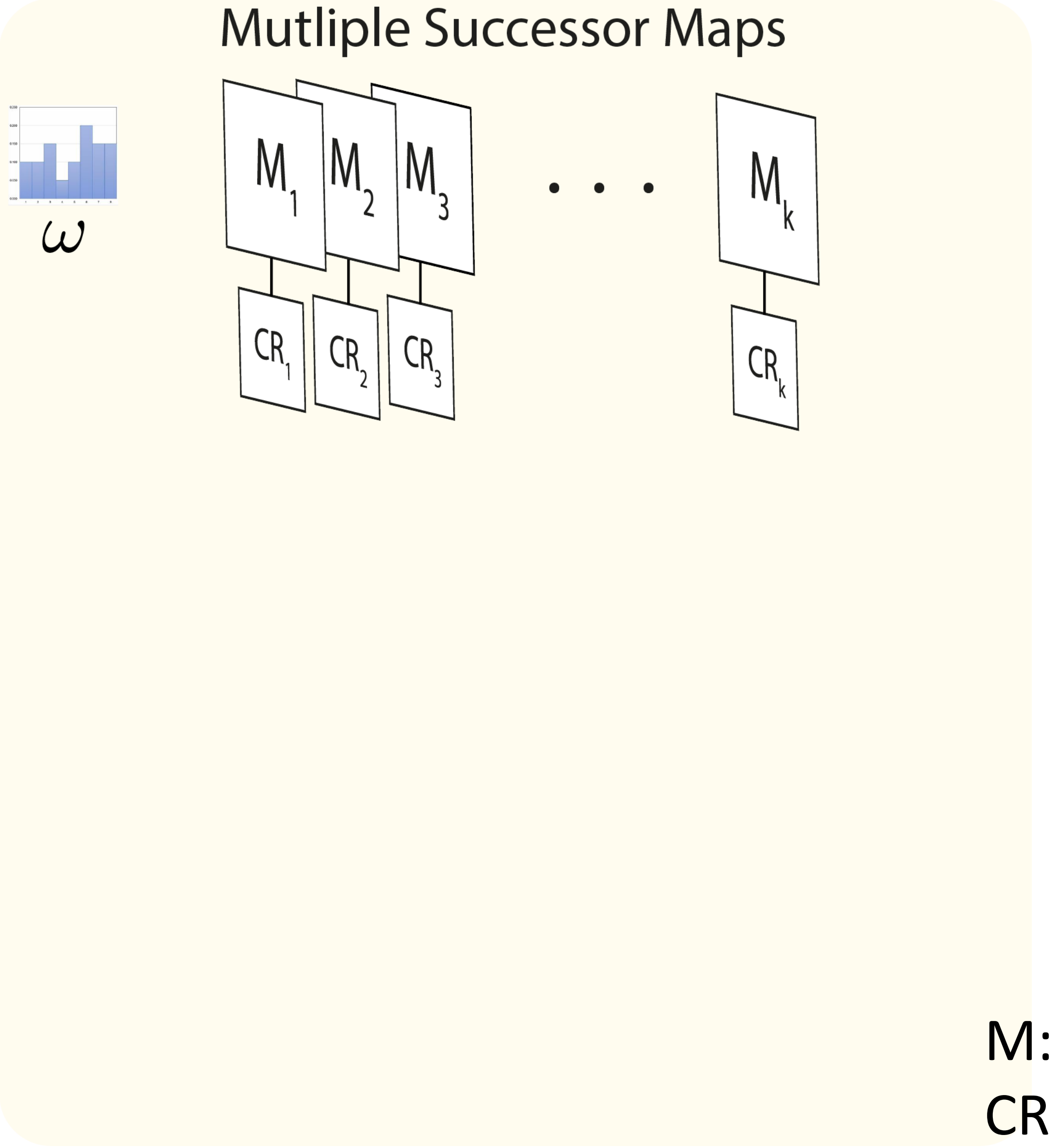


Generative model over reward functions



Dirichlet Process mixture model of kernel- smoothed rewards

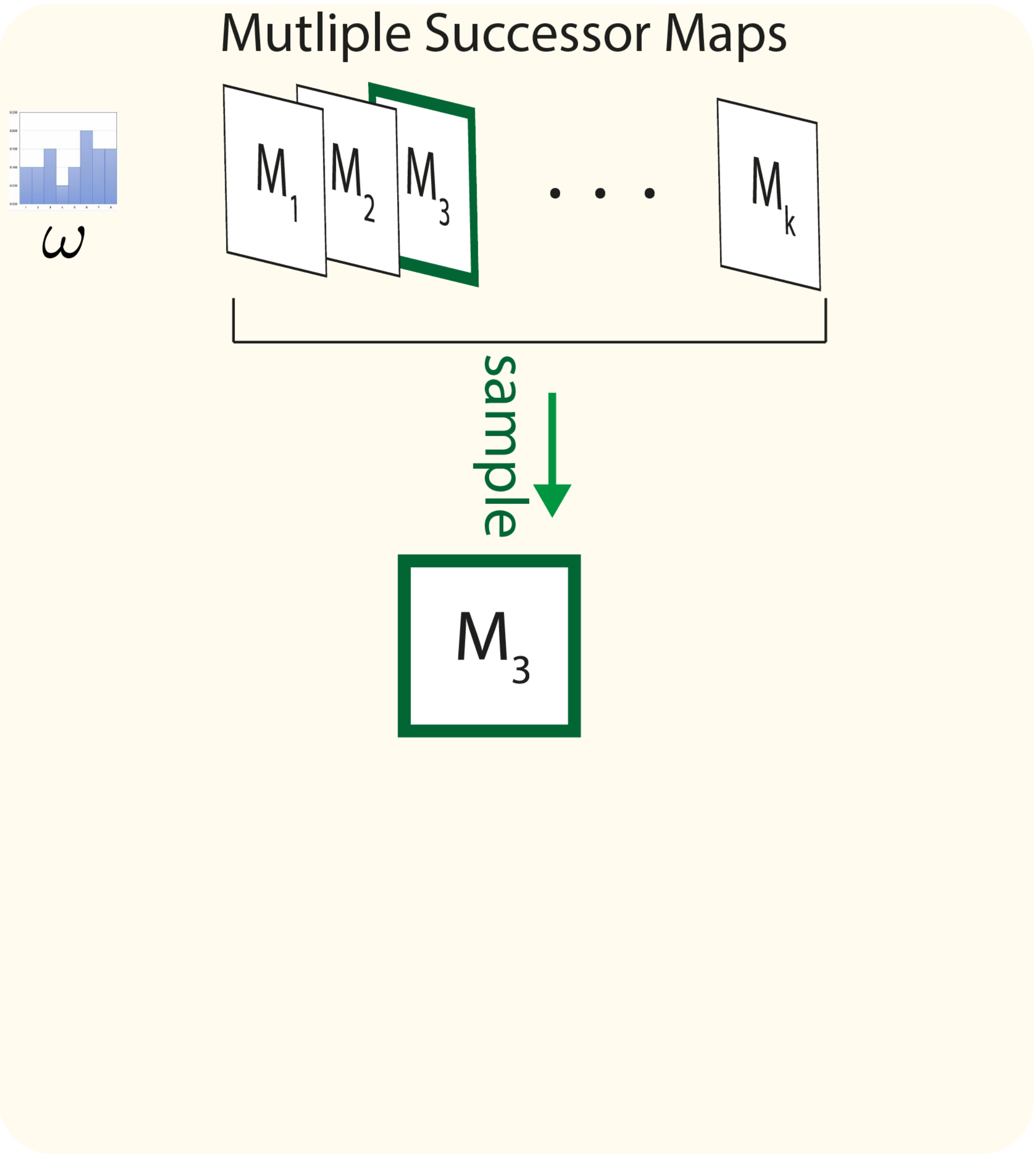
Bayesian Successor Representation (BSR)



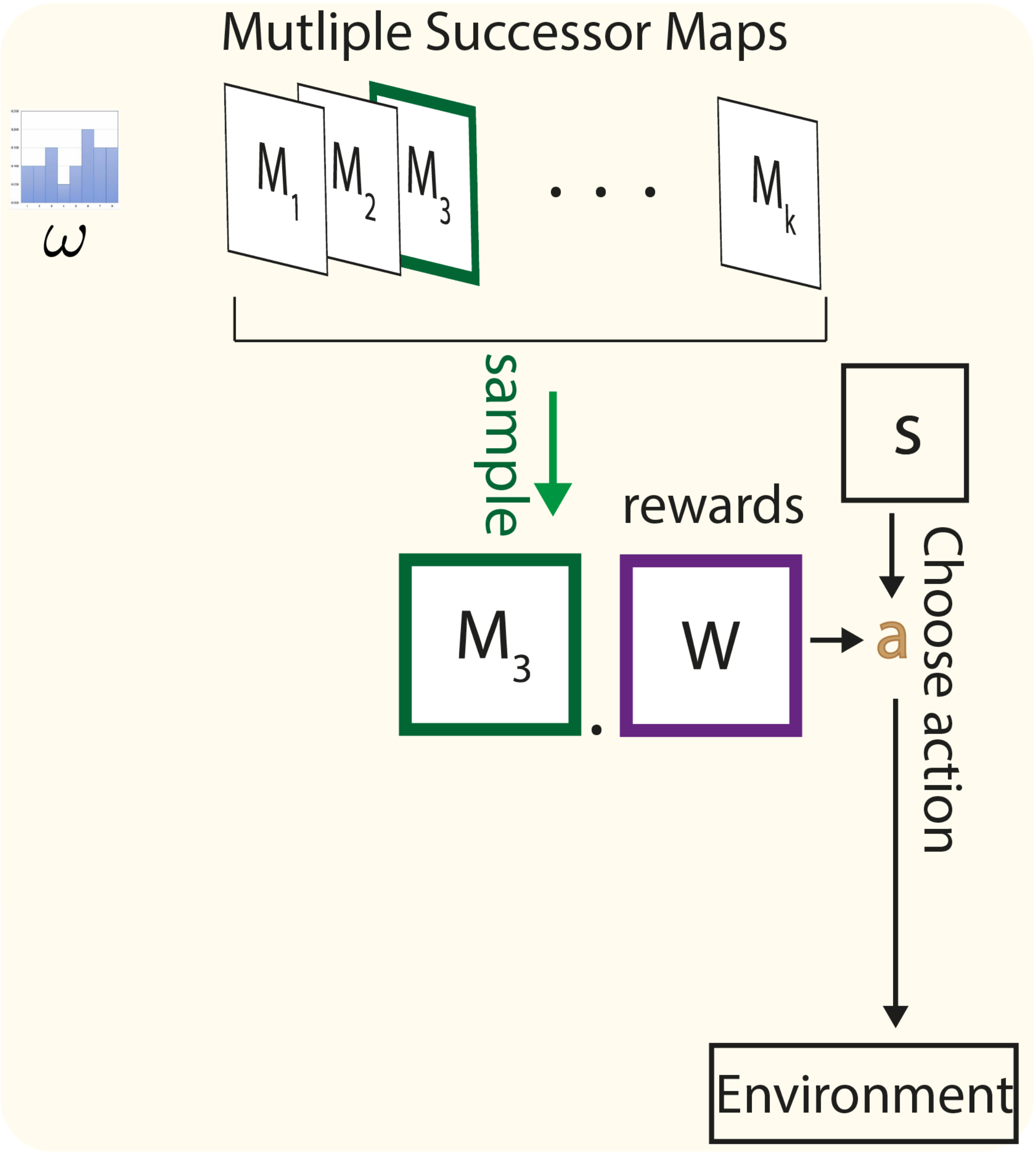
M: Successor Representation

CR: Convolved reward map

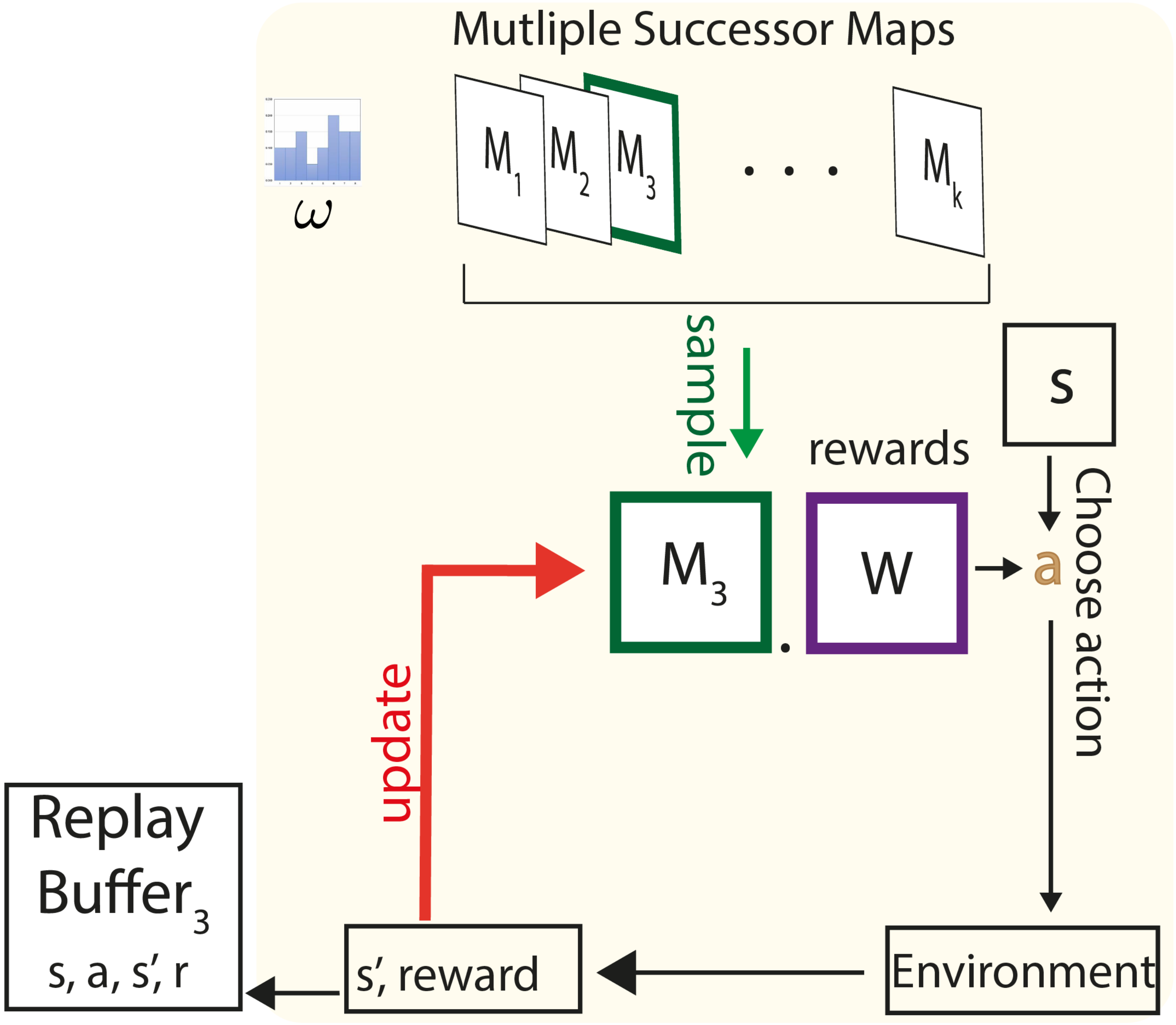
Bayesian Successor Representation (BSR)



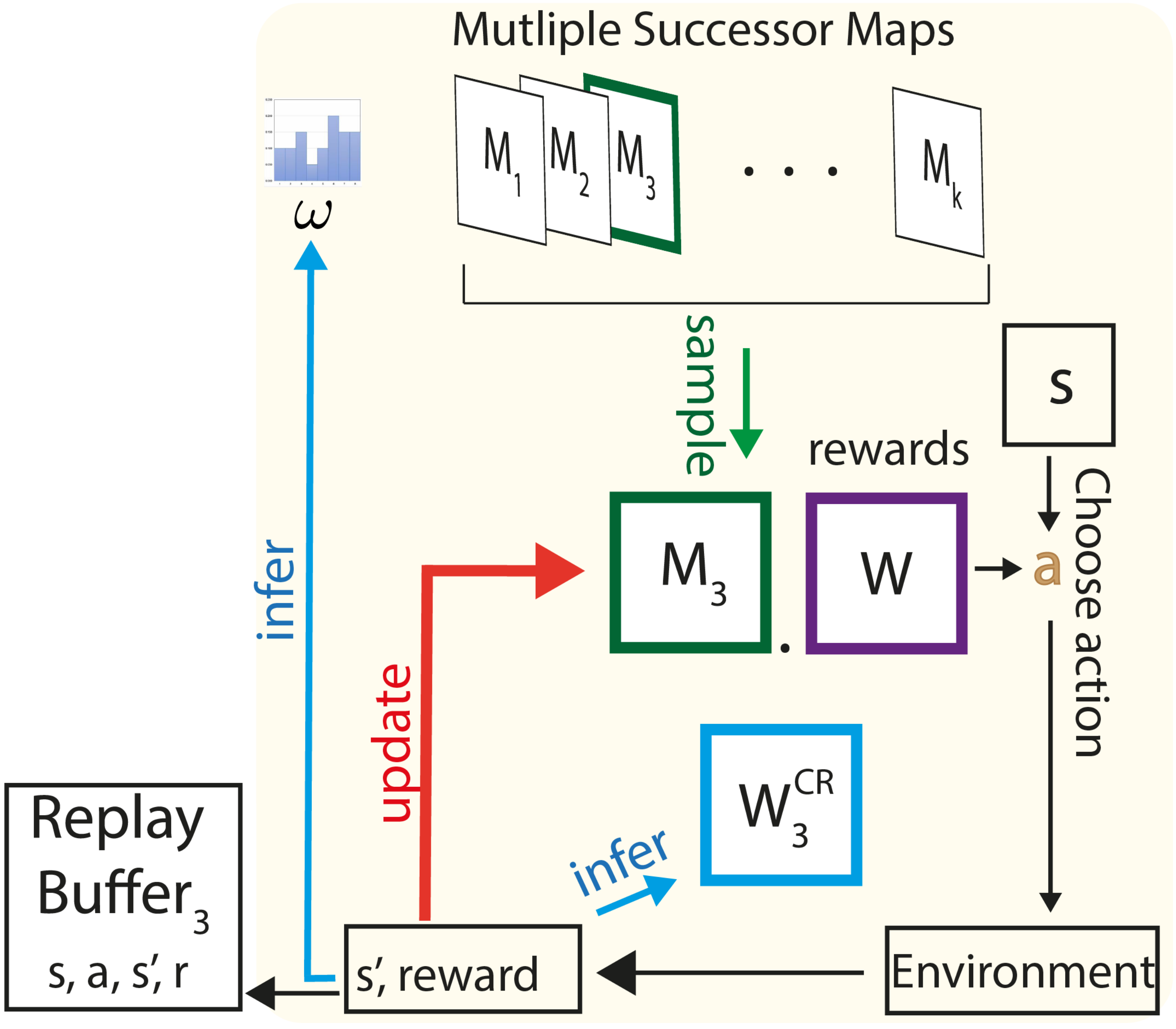
Bayesian Successor Representation (BSR)



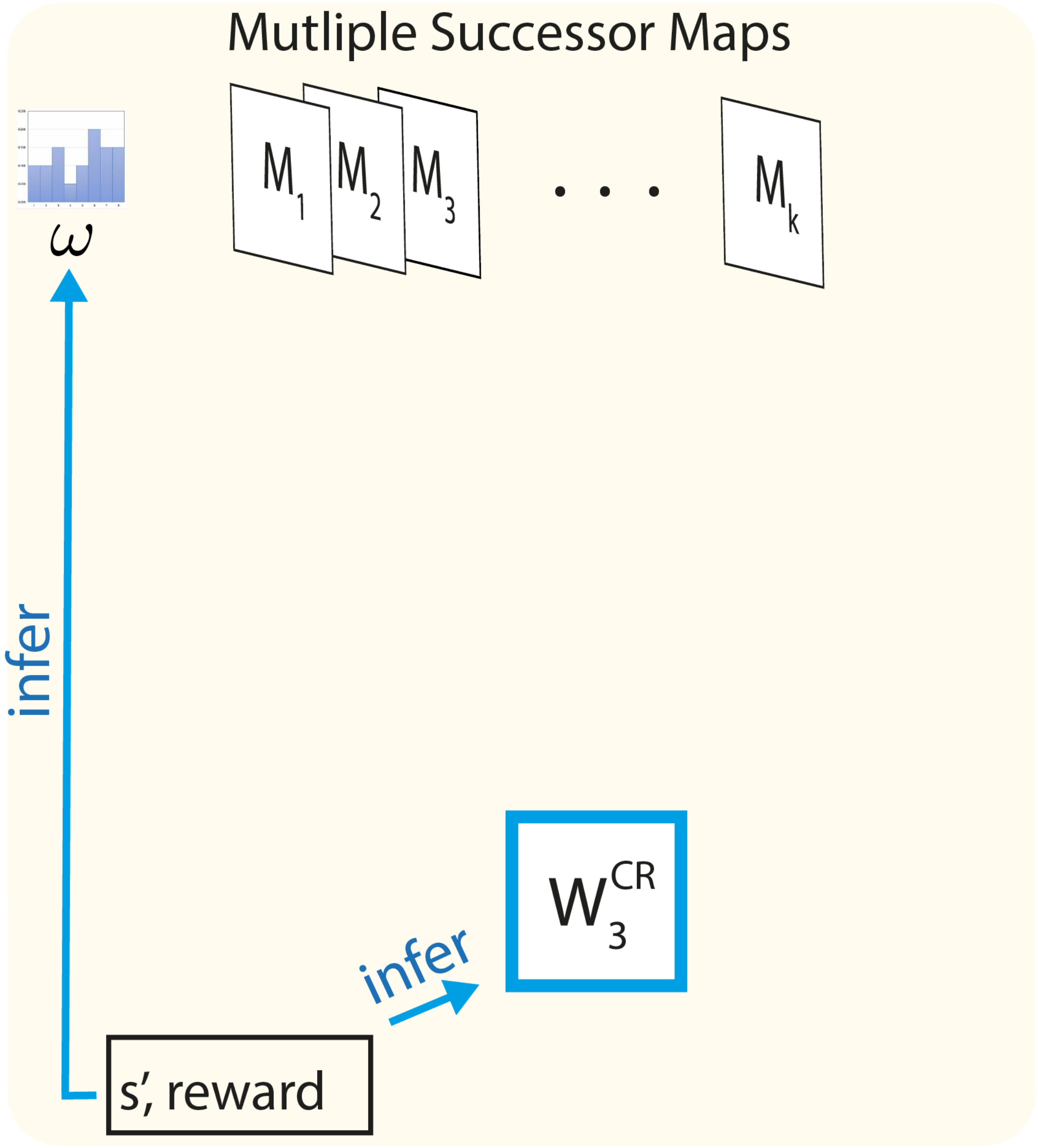
Bayesian Successor Representation (BSR)



Bayesian Successor Representation (BSR)

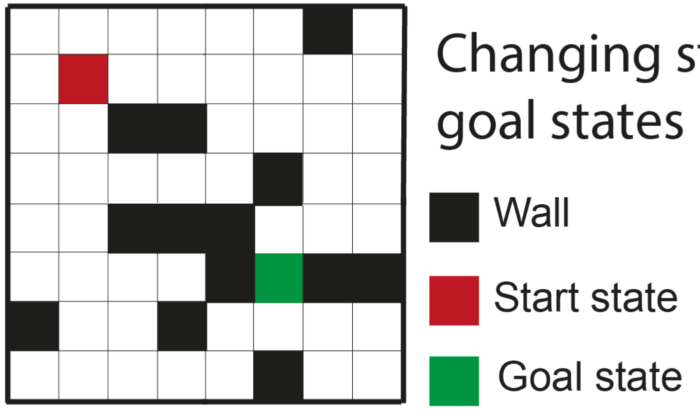


Bayesian Successor Representation (BSR)

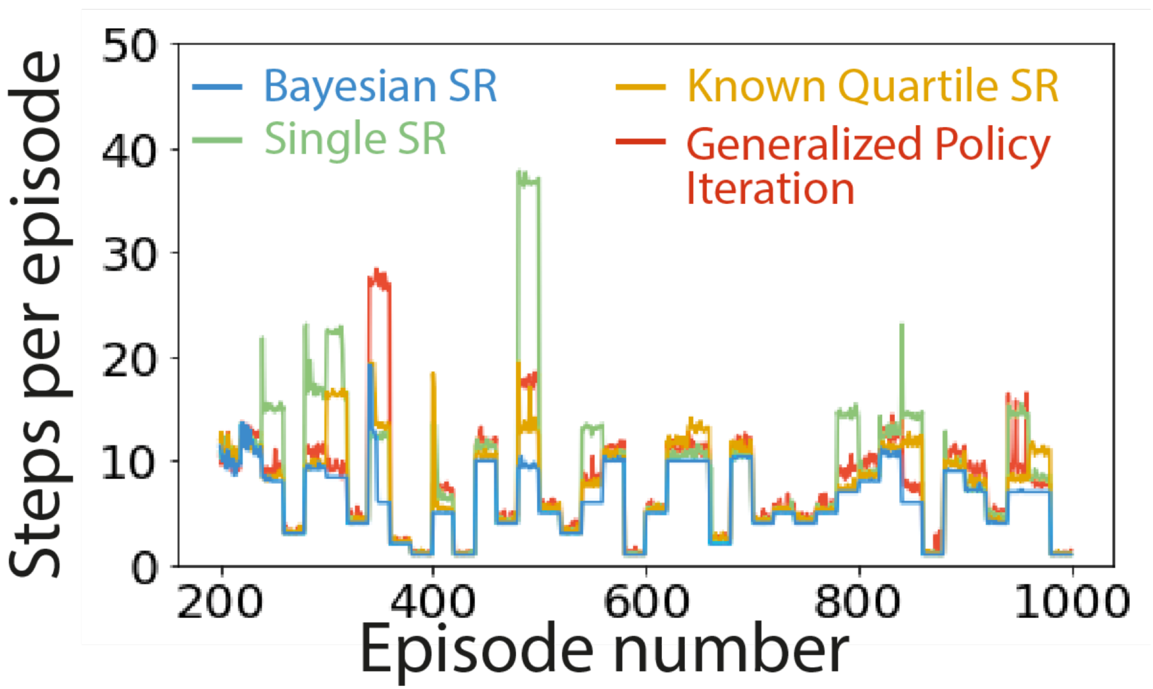


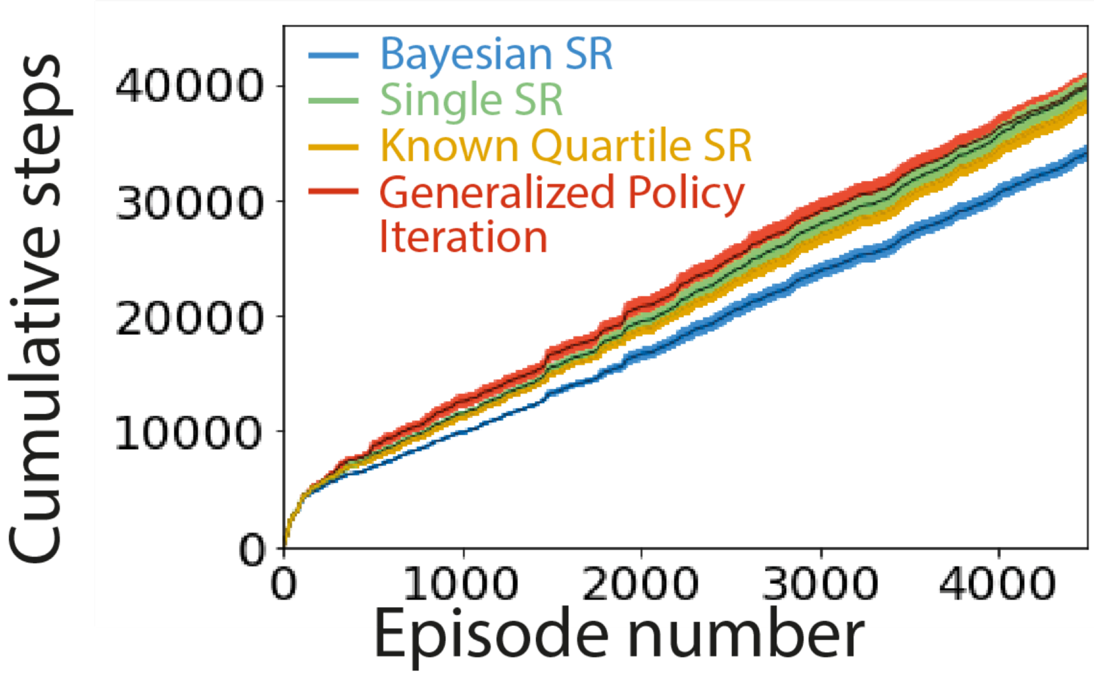
Results

Dynamic maze navigation



Changing start and goal states

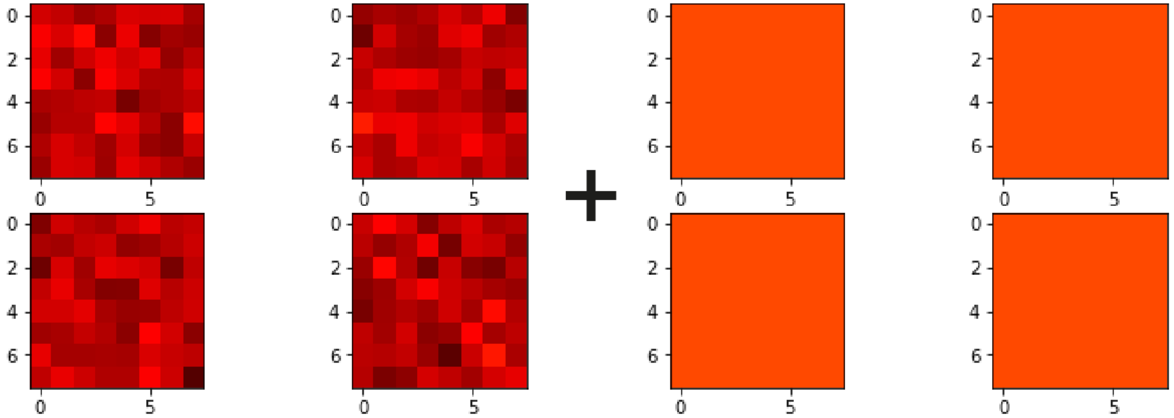




Barreto et al. 2017
NeurIPS

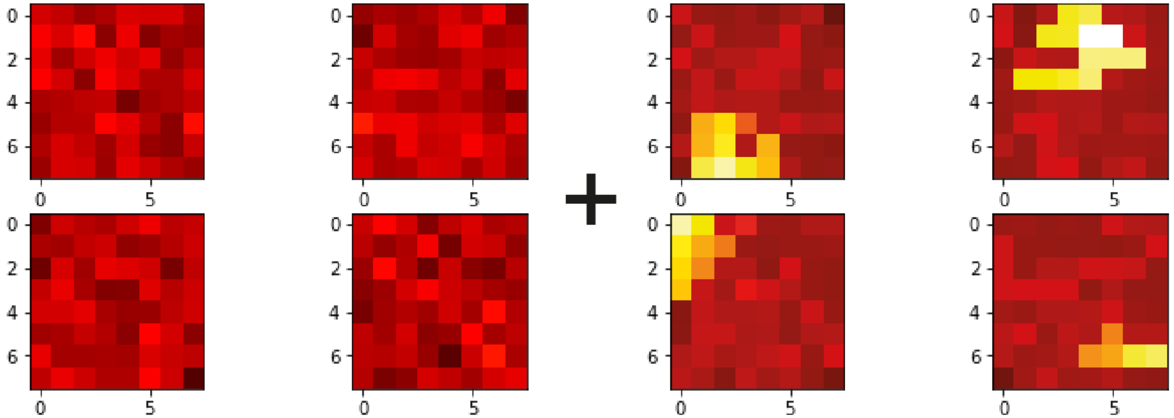
Multi-task exploration bonus by offsetting the reward belief vector **w**

w



UCB inspired constant offset

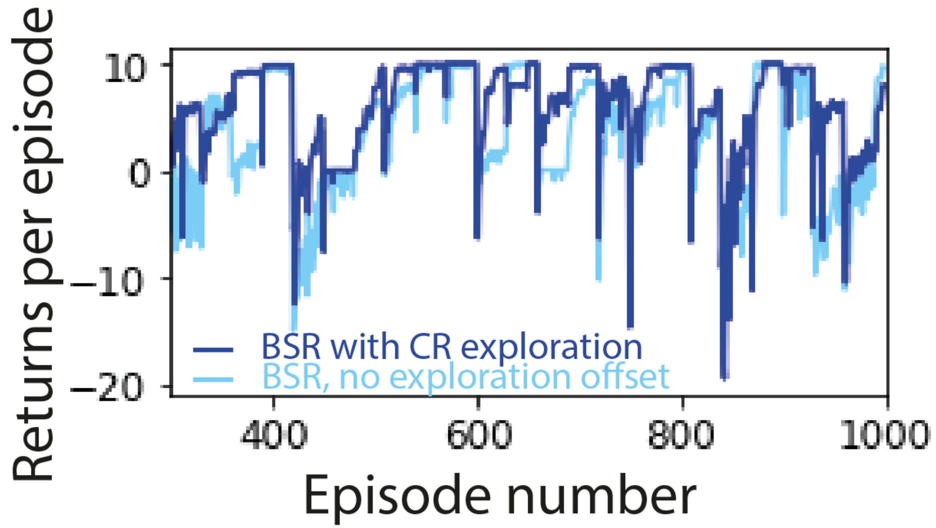
w

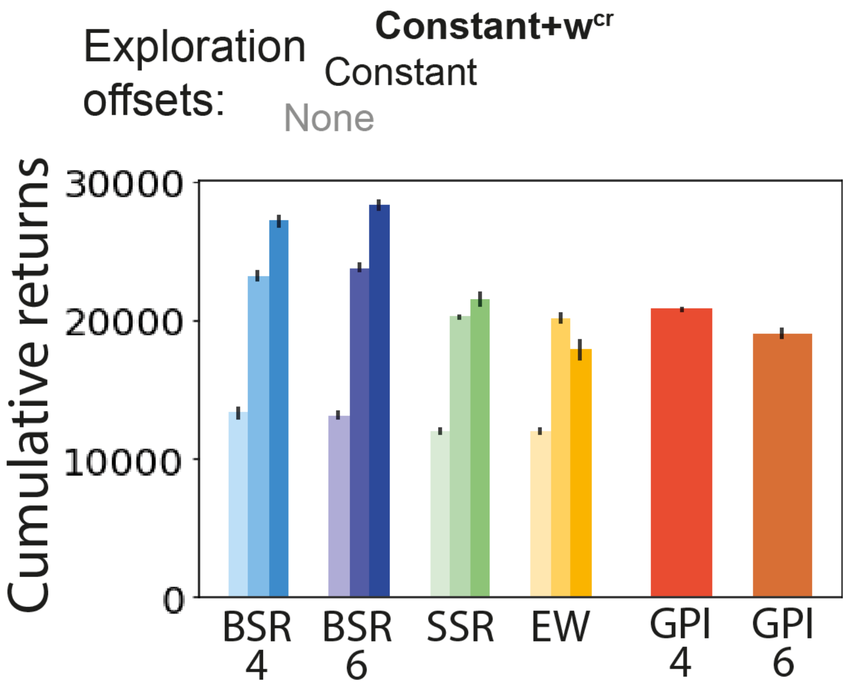


Offset using CR maps, acting as priors for rewards

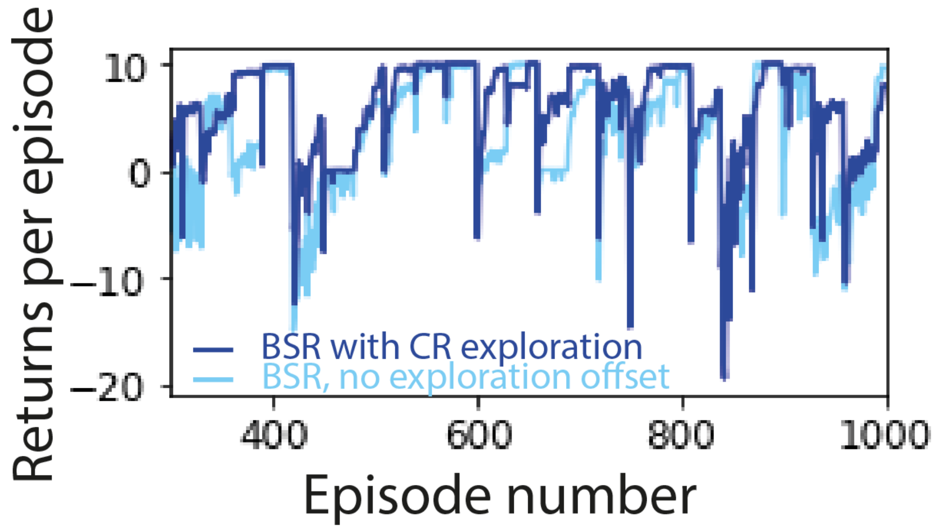
Auer 2002
JMLR

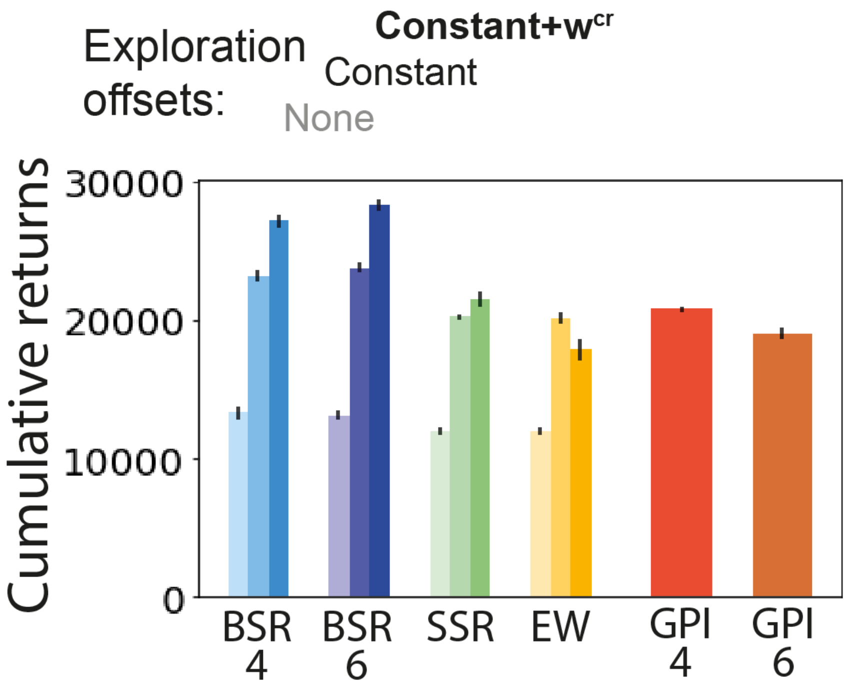
Results

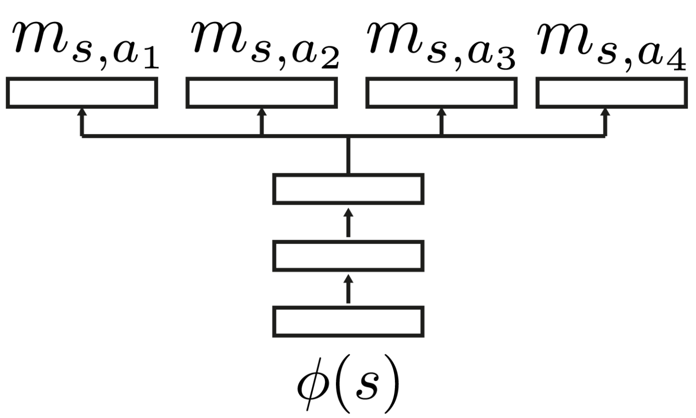

10
0
−10
−20
400
600
800
1000
Returns per episode
Episode number
BSR with CR exploration
BSR, no exploration offset

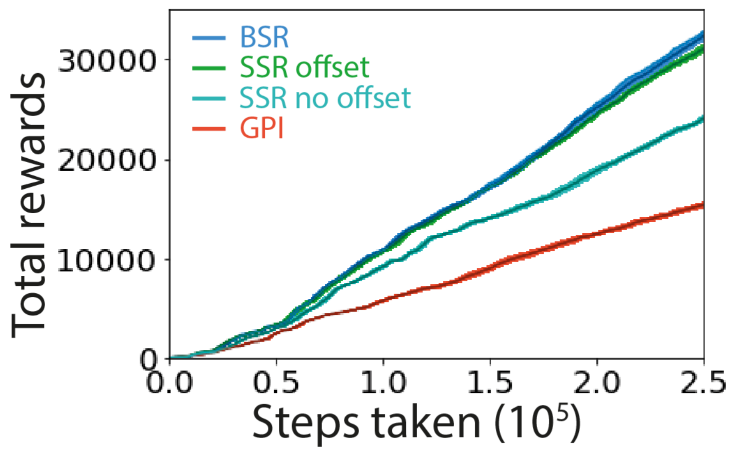

Exploration offsets:
Constant+w^{cr}
Constant
None
Cumulative returns
30000
20000
10000
0
BSR 4
BSR 6
SSR
EW
GPI 4
GPI 6

Results


Returns per episode
10
0
−10
−20
400
600
800
1000
Episode number
BSR with CR exploration
BSR, no exploration offset


Exploration offsets:
Constant+w^cr
Constant
None
Cumulative returns
30000
20000
10000
0
BSR 4
BSR 6
SSR
EW
GPI 4
GPI 6

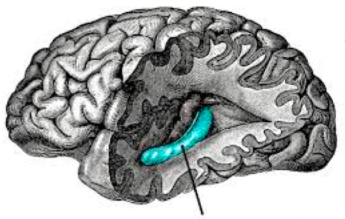

m_{s,a_1} m_{s,a_2} m_{s,a_3} m_{s,a_4}
$\phi(s)$


BSR
SSR offset
SSR no offset
GPI
Total rewards
30000
20000
10000
0
0.0
0.5
1.0
1.5
2.0
2.5
Steps taken (10^5)

Results

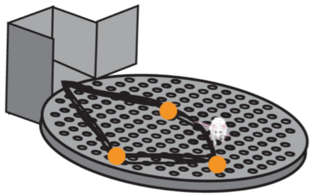
Data

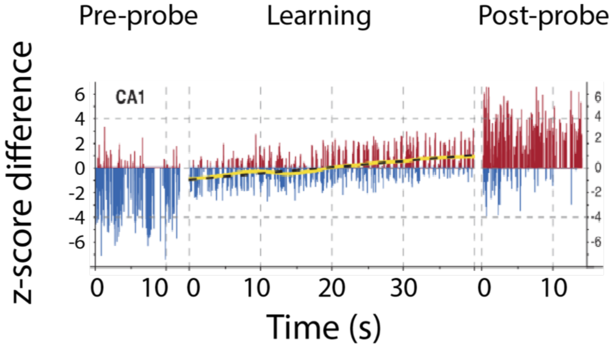
BSR

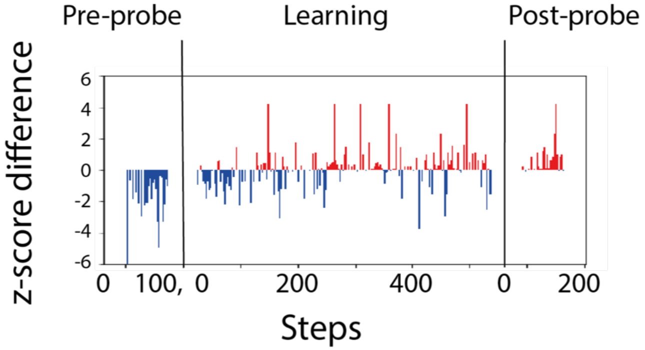


Blum and Abbot 1996
Levy et al. 2005
Stachenfeld et al. 2017

Finding changing set of rewards

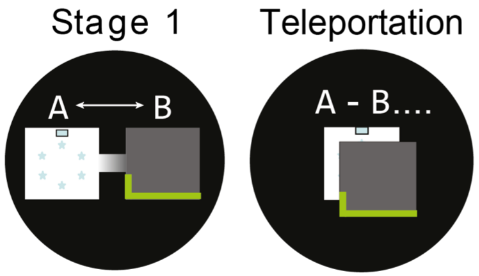


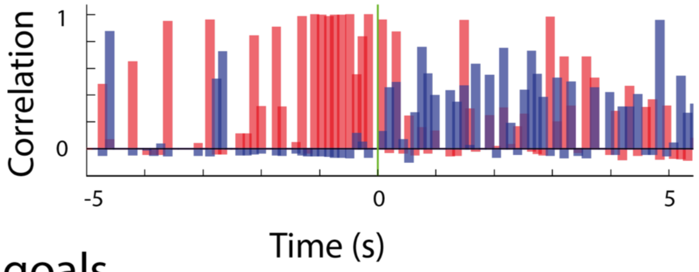


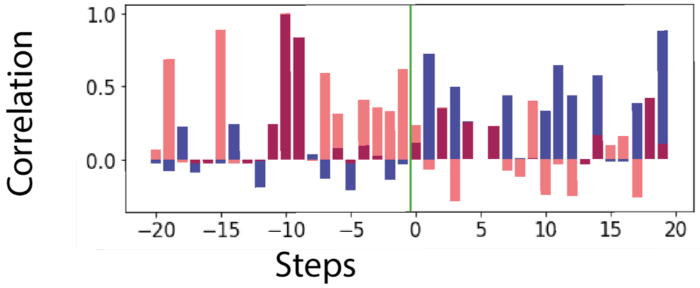


Teleportation

Boccara et al. 2019
Science

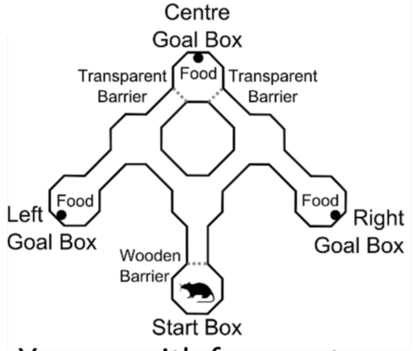




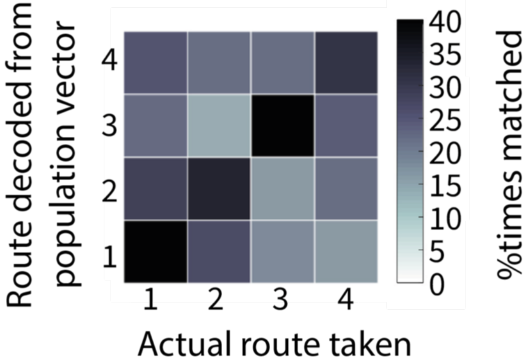


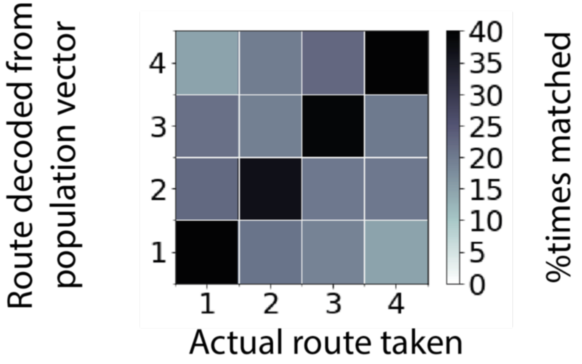
Jezek et al. 2019
Nature

Navigation with changing goals and barriers



Y-maze with four routes





Grieves et al. 2016
Elife

Thank you!

arXiv:1906.07663

Transfer and Multi-task learning
Poster #52
10:45 AM - 12:45 PM