


## “Living Penalty”

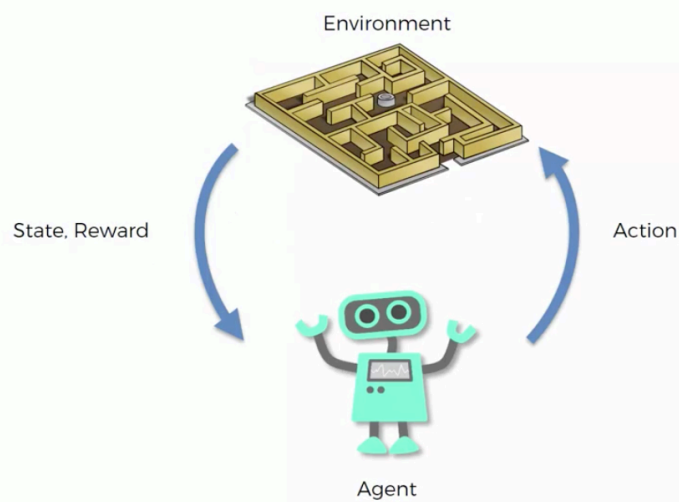
## “Living Penalty”

$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$

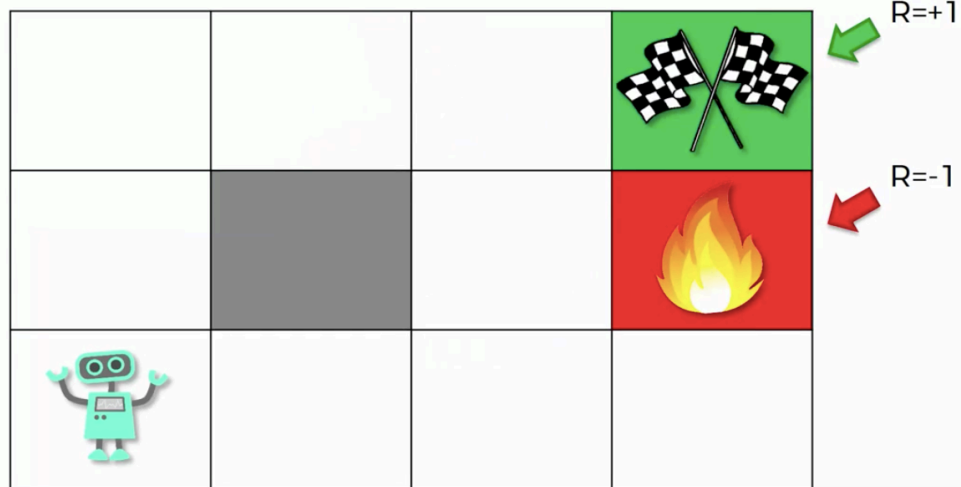
## “Living Penalty”


$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$

## “Living Penalty”



## “Living Penalty”

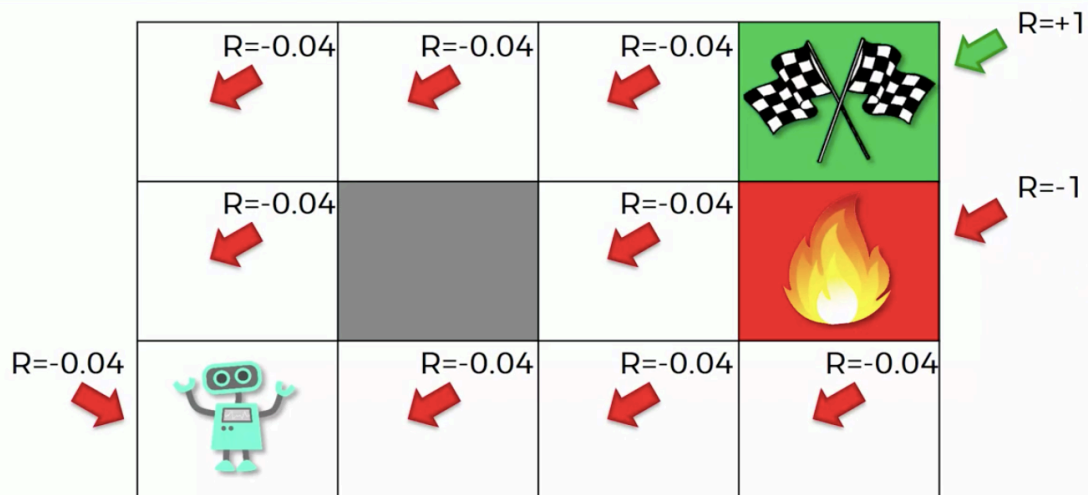


Artificial Intelligence A-Z

© SuperDataScience

So far, we gave the agent reward at the end but we want to give it reward throughout the way as well.

## “Living Penalty”



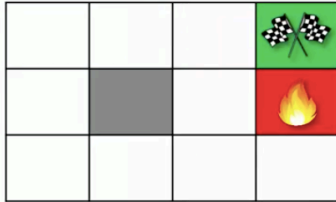
Artificial Intelligence A-Z

© SuperDataScience

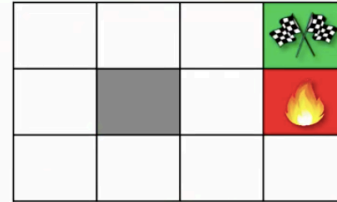
It's important to remember that an agent gets this negative reward only when it enters tile and it doesn't get this at the very start. With this strategy, agent wants to get the reward as soon as possible.

## “Living Penalty”

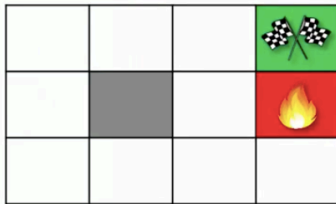
$R(s)=0$



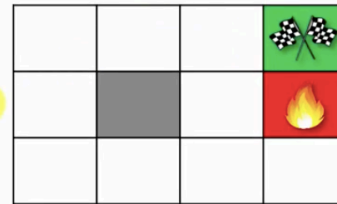
$R(s)=-0.04$



$R(s)=-0.5$



$R(s)=-2.0$



Artificial Intelligence A-Z

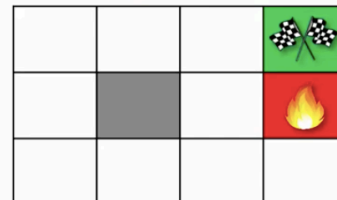
© SuperDataScience

## “Living Penalty”

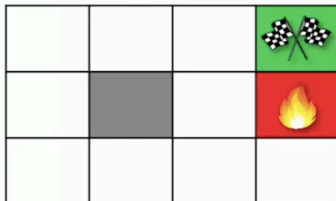
$R(s)=0$



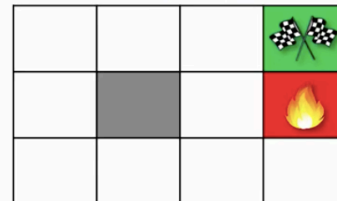
$R(s)=-0.04$



$R(s)=-0.5$



$R(s)=-2.0$

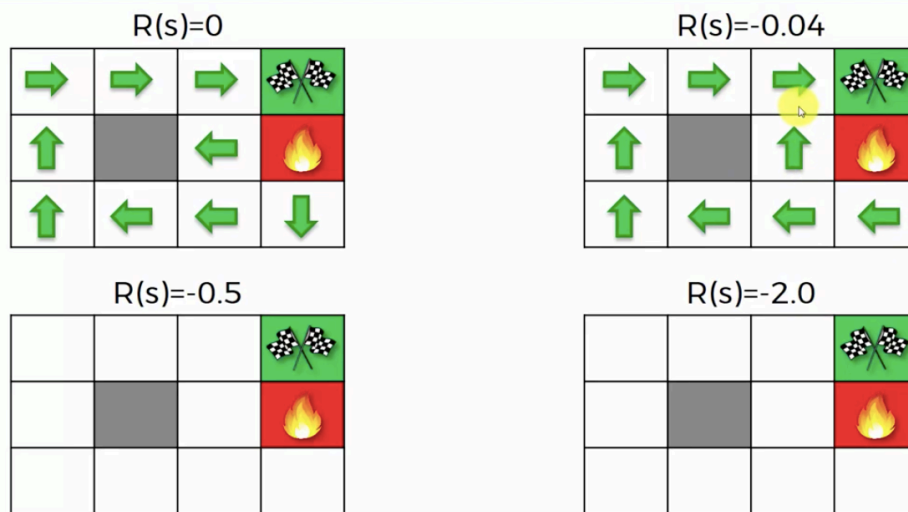


Artificial Intelligence A-Z

© SuperDataScience

Here the reward or the living penalty is equal to 0 which is our original policy.

## “Living Penalty”

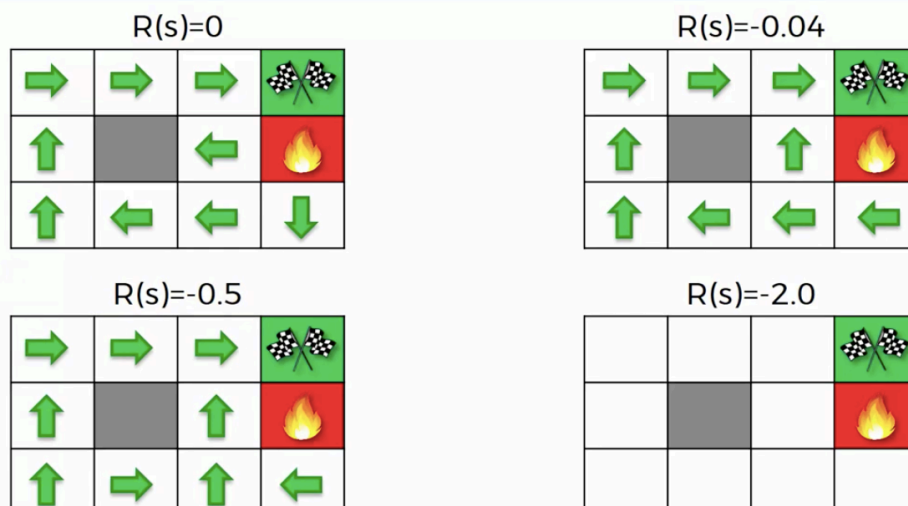


Artificial Intelligence A-Z

© SuperDataScience

If we add a negative reward of 0.4, we can see instantly that the two arrows around the fire pit has been changed. Which it means the agent doesn't want jump to the wall and it's willing to take the risk of going to the fire pit but it goes from the right path (not 100% but more likely). The reason is because of the living penalty which the agent assumes it's going to be worse for it, if it goes from the left path.

## “Living Penalty”

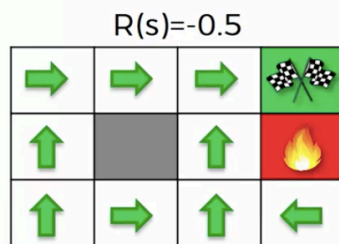
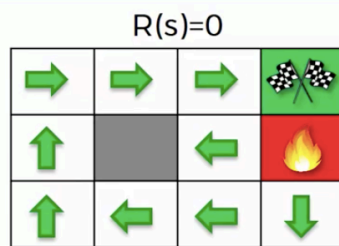


Artificial Intelligence A-Z

© SuperDataScience

If we increase the living penalty, the agent's willingness for going from the right path will increase.

## “Living Penalty”



In here, the agent is even willing go to the fire pit in order to reach to the positive reward. Since the living penalty is even twice is much the fire pit (negative-wise)