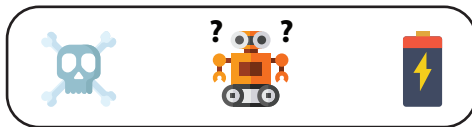
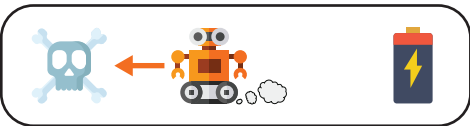


Observe the environment

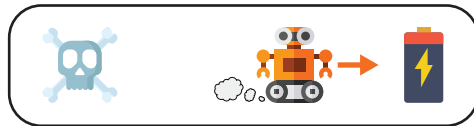


Choice A



Select an action using the policy matrix

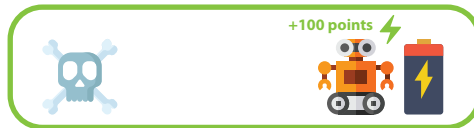
Choice B



Observe the consequences



Punishment



Reward

Update the policy matrix

Actions			Actions	
A	B		A	B
-1	+2	Bellman equation	+10	+2
0	+1		-4	+5
-3	-9		-1	0
t=n			t=n+1	

Iterate ...